



مركز البحوث والدراسات

# تطوير المقاييس النظرية والتطبيق

تأليف:  
روبرت ف. ديفيليس



ترجمة

أ.د. طارق عطية عبدالرحمن

د. سعد بن سعيد القحطاني

راجع الترجمة

أ.د. عبدالرحمن بن سليمان العنقري

بسم الله الرحمن الرحيم



مركز البحوث والدراسات

تطوير

المقاييس

النظرية والتطبيق

تأليف

روبرت ف ديفيليس

ترجمة

د سعد بن سعيد القحطاني      أد طارق عطية عبدالرحمن

راجع الترجمة

أد عبدالرحمن بن سليمان العنقري

1442هـ - 2021م

مكتبة الحير الإلكتروني

مكتبة العرب الحصرية

## بطاقة الفهرسة

③ معهد الإدارة العامة، 1442هـ

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

ديفيليس، روبرت ف

تطوير المقاييس: النظرية والتطبيق / روبرت ف ديفيليس؛ سعد

بن سعيد القحطاني؛ طارق عطية عبدالرحمن؛ عبدالرحمن

سليمان العنقري - الرياض، 1442هـ

344 ص؛ 17 سم × 24 سم

ردمك: 7 - 29 - 8276 - 603 - 978

1- الموازين والمقاييس أ القحطاني، سعد بن سعيد (مترجم) ب

عبدالرحمن، طارق عطية (مترجم) ج العنقري، عبدالرحمن

سليمان (مراجع) د العنوان

ديوي 3891

1442/3085

رقم الإيداع: 1442/3085

ردمك: 7 - 29 - 8276 - 603 - 978

هذه ترجمة لكتاب:

**SCALE DEVELOPMENT**

**Theory and Applications**

**By**

**Robert F DeVellis**

## قائمة المحتويات المختصرة

الصفحة	الموضوعات
15	مقدمة
18	شكر وتقدير
18	نبذة عن المؤلف
21	الفصل الأول: نظرة عامة
47	الفصل الثاني: فهم المتغير الكامن
67	الفصل الثالث: الثبات (الموثوقية)
121	الفصل الرابع: الصدق
147	الفصل الخامس: قواعد إرشادية في تطوير المقياس
209	الفصل السادس: التحليل العاملي

273	الفصل السابع: نظرة عامة على نظرية الاستجابة للبند (IRT)
307	الفصل الثامن: القياس في سياق البحث الأوسع
325	المراجع

## قائمة المحتويات التفصيلية

الصفحة	الموضوعات
15	مقدمة
18	نبذة عن المؤلف
21	الفصل الأول: نظرة عامة
23	المنظورات العامة بشأن القياس
24	الأصول التاريخية للقياس في العلوم الاجتماعية
24	أمثلة مبكرة
27	ظهور الأساليب الإحصائية ودور الاختبارات العقلية
27	دور علم النفس الفسيولوجي
29	التطورات اللاحقة في القياس
29	تطور المفاهيم الأساسية
29	تطور الاختبارات العقلية

30	تقييم المرض العقلي
35	توسيع نطاق القياس النفسي
36	دور القياس في العلوم الاجتماعية
36	علاقة النظرية بالقياس
38	المقاييس النظرية وغير النظرية
39	أنواع المقاييس
42	جميع المقاييس ليست متساوية
44	تكلفة سوء أو ضعف القياس
45	الملخص والمراجعة
46	تمارين
47	
	<b>الفصل الثاني: فهم المتغير الكامن</b>
47	المفاهيم البنائية (constructs) مقابل المقاييس
49	المتغير الكامن كسبب مفترض لقيم البند
50	مخططات المسار
51	الاصطلاحات التخطيطية



52	مخططات المسار في تطوير المقياس
55	مزيد من التوسع في نموذج القياس
55	افتراضات القياس الكلاسيكية (التقليدية)
56	الاختبارات المتكافئة (المتوازية)
60	نماذج بديلة
63	اختيار نموذج سببي
65	تمارين
65	ملاحظات
67	الفصل الثالث: الثبات (الموثوقية)
68	الطرق القائمة على تحليل التباين
70	البنود المتصلة «المستمرة» مقابل البنود ثنائية الاستجابة
71	الاتساق الداخلي
72	معامل «ألفا» Alpha
73	مصفوفة التغاير
74	مصفوفات التغاير للمقاييس متعددة البنود
75	معامل «ألفا» ومصفوفة التغاير

81	صيغة أو معادلة بديلة لمعامل «ألفا»
84	
	نقد معامل «ألفا»
90	معالجة محددات وقيود معامل «ألفا»
92	معامل أوميغا ( $\omega$ )
95	الثبات القائم على الارتباطات بين درجات المقياس
95	ثبات النماذج البديلة
96	ثبات التجزئة النصفية
101	اتفاق المُقيّمين inter-rater agreement
103	الاستقرار الزمني
107	ثبات درجات التعبير
114	الثبات والقوة الإحصائية
115	نظرية التعميم
118	ملخص
119	تمارين
120	ملاحظات

121	الفصل الرابع: الصدق
122	صدق المحتوى أو المضمون
124	نطاق المتغير والآثار المترتبة على صدق المحتوى
132	الصدق المرتبط بمحك (معياري)
133	الصدق المرتبط بمحك مقابل الدقة
135	الصدق البنائي (صدق التكوين الفرضي)
136	التمييز بين الصدق البنائي والصدق المرتبط بمحك
138	التخفيف (التضعيف) Attenuation
140	إلى أي مدى يجب أن تكون الارتباطات قوية لإظهار الصدق البنائي؟
140	matrix مصفوفة الصفات المتعددة - الأساليب المتعددة Multitrait-Multimethod
143	ماذا عن الصدق الظاهري؟
145	تمارين
147	الفصل الخامس: قواعد إرشادية في تطوير المقياس
147	الخطوة الأولى: حدد بوضوح ماذا تريد قياسه
147	النظرية كوسيلة مساعدة على تحقيق الوضوح

148	الخصوصية كوسيلة مساعدة على الوضوح
150	توضيح ما يجب تضمينه في القياس
151	الخطوة الثانية: توليد مجموعة من البنود
152	اختر البنود التي تعكس الغرض من المقياس
153	التكرار
157	عدد البنود
158	بدء عملية كتابة البنود
158	خصائص البنود الجيدة والبنود السيئة
162	البنود ذات الصياغة الإيجابية والبنود ذات الصياغة السلبية
163	الخلاصة
164	الخطوة الثالثة: تحديد شكل أو تنسيق القياس
164	مقياس ثرستون (Thurstone)
166	مقياس جتمان (Guttman)
168	المقاييس ذات البنود متساوية الترجيح أو الوزن
168	كم عدد فئات الاستجابة؟
174	أنواع محددة (معينة) من أشكال وتنسيقات الاستجابة

174	مقياس ليكرت
178	القياس التفاضلي الدلالي (تباين الدلالات اللفظية)
179	التناظر البصري
181	أشكال الاستجابة العددية والعمليات العصبية الأساسية
182	الخيارات الثنائية
183	الأطر الزمنية للبند
185	الخطوة الرابعة: مراجعة تجمع البنود الأولية من قبل الخبراء
187	الخطوة الخامسة: النظر في إدراج وتضمين البنود التي تحقق الصدق
188	الخطوة السادسة: تطبيق البنود على عينة تطوير المقياس
191	الخطوة السابعة: تقييم البنود
191	الفحص المبدئي لأداء البنود
192	عكس تسجيل الدرجات
194	ارتباطات البند بالمقياس
195	تباينات البنود
196	متوسطات البنود
197	البعدية (الأبعاد)

197	الثبات (الموثوقية)
201	الخطوة الثامنة: تحقيق الحد الأمثل لطول المقياس
201	تأثير طول المقياس على الثبات
201	آثار استبعاد وإسقاط البنود السيئة
203	ضبط وإصلاح طول المقياس
204	تقسيم العينات
206	تمارين
207	ملاحظات
209	الفصل السادس: التحليل العاملي
211	نظرة عامة حول التحليل العاملي
212	أمثلة على الطرق المماثلة للمفاهيم التحليلية للعوامل
212	مثال رقم ١
214	مثال رقم ٢
215	أوجه القصور في هذه الطرق
218	الوصف المفاهيمي للتحليل العاملي
219	استخلاص العوامل

219	العامل الأول
222	العوامل اللاحقة
223	تحديد عدد العوامل التي يجب استخلاصها
230	تدوير العوامل
231	التشبيه الأول لتدوير العوامل
233	التشبيه الثاني لتدوير العوامل
236	التشبيه الثالث لتدوير العوامل
242	التدوير المتعامد مقابل التدوير المائل
246	اختيار نوع التدوير
248	النماذج العاملية الثنائية والهرمية
257	تفسير العوامل
259	طريقة المكونات الرئيسية مقابل العوامل المشتركة
259	هل هما نفس الشيء أم مختلفان؟
263	التحليل العاملي التوكيدي Analysis Confirmatory Factor
266	استخدام التحليل العاملي في تطوير المقاييس
270	حجم العينة

271	الخلاصة
271	تمارين
273	الفصل السابع: نظرة عامة على نظرية الاستجابة للبند (IRT)
277	صعوبة البند Item Difficulty
279	تمييز البند Item Discrimination
280	التخمين، أو الإيجابيات الزائفة Guessing, or False Positives
282	ITEM-CHARACTERISTIC CURVES المنحنيات المميزة للبند (ICC)
287	تطبيق نظرية الاستجابة للبند IRT على البنود متعددة الاستجابة
295	ثيتا والاختبار التكيفي المحوسب Theta and Computerized Adaptive Testing (CAT)
299	تعقيدات نظرية الاستجابة للبند IRT
302	الخلاصة
305	تمارين
307	الفصل الثامن: القياس في السياق البحثي الأوسع
307	قبل تطوير المقياس
307	ابحث عن الأدوات الحالية



309	عرض المفهوم البنائي في سياق المجتمع محل الاهتمام
312	اتخاذ القرار بشأن طريقة تطبيق المقياس
312	الأخذ في الاعتبار المقياس في سياق مقاييس أو إجراءات أخرى
313	بعد تطبيق المقياس
314	قضايا تحليلية
314	قضايا التفسير
315	قابلية التعميم
316	أفكار ختامية
316	القياس الصغير والقياس الكبير
316	الزوارق والسفن البحرية
318	القياس «الزوارق» والقياس «سفن الرحلات البحرية»
320	الآثار العملية للقياس الصغير مقابل القياس الكبير
322	تذكّر، القياس مسألة هامة
323	تمارين
325	المراجع

## مقدمة:

تطوير المقاييس مجال يتطور بسرعة مع ظهور مناهج جديدة واكتساب المزيد من القوة وقد نشأت هذه التغييرات من مجموعة متنوعة من المصادر، بما في ذلك الجهود الكبيرة لتوفير المزيد من أدوات القياس الموحدة بشكل أكبر عبر مجموعة واسعة من السياقات فمن ناحية، تتسم هذه التغييرات بالأهمية والإثارة ومن ناحية أخرى، فإنها تجعل مراجعة وتنقيح هذا الكتاب تحدياً إلى حد كبير وكما هو الحال مع الإصدارات أو الطبقات السابقة، فإن هدفي هو تقديم معلومات دقيقة أو معقدة بطريقة يمكن لمجموعة من الباحثين الوصول إليها وفي الوقت نفسه، ومع كل مراجعة وتنقيح، حاولت جاهداً تحديث المادة العلمية لتعكس الاتجاهات الحالية في هذا المجال ويتعارض هذان الهدفان في بعض الأحيان تعارضاً تاماً، حيث إن بعض التطورات الجديدة تتطلب على مفاهيم ومنهجيات متقدمة ومتخصصة نوعاً ما وقد سعيت إلى حل هذه المعضلة عن طريق توفير سياق للتطورات الجديدة، وعن طريق التمييز بشكل واضح بين أنشطة القياس المحدودة والضيقة وأنشطة القياس الأوسع نطاقاً والنهج والمداخل التقنية التي يمكن أن تناسب كلاً منها وبطبيعة الحال، فقد اقتصررت مناقشتي لهذه القضايا والمسائل على سياق نطاق الكتاب وهدفه الذي يعد بمثابة نظرة عامة على النهج السيكمترية بدلاً من المراجعة الشاملة

وكانت المحصلة النهائية هي أن كل فصل يحتوي على مادة جديدة مضافة إليه في الفصل الأول، نظرة عامة، يتضمن مراجعتي واستعراضاتي لتاريخ القياس الآن ملخصاً لكيفية تصنيف الأمراض العقلية من العصور القديمة إلى الوقت الحاضر، مع التركيز على الأنشطة الأكثر حداثة ويكشف هذا النقاش عن الفرص الضائعة وغيرها من الفرص التي تم استغلالها

في الفصل الثاني، فهم المتغير الكامن، قمت بإضافة قسم يتناول بشكل مباشر أكبر الآثار المترتبة على اختيار النموذج السببي الذي يربط البنود بمتغيراتها الكامنة وقد لاحظت أن هذا الاعتبار المهم والواضح على ما يبدو، كثيراً ما يتم تجاهله وإغفاله وبعض النماذج التي تربط المتغيرات بمؤشراتها تكون أكثر ملاءمةً لُنهج القياس المباشرة أكثر من غيرها وبناءً على ذلك، أعتقد أن هذه النقطة تستدعي وتحتاج مزيداً من التركيز

وتمثل التغييرات في الفصل الثالث، الثبات، التطور السريع في التفكير في بعض مجالات القياس وقد ظهرت في السنوات الأخيرة مجموعة كبيرة من الأعمال التي تشكك في استخدام معامل «كرونباخ ألفا» وعلى الرغم من أنني قد أشرت إلى هذه المنظورات الجديدة في الطبعة السابقة، إلا أنني قد شعرت بالحاجة إلى التوسع في تلك المناقشة في الطبعة الحالية من الكتاب وهذا يتضمن تسليط الضوء على بعض الانتقادات والظروف التي تنطبق عليها، فضلاً عن مناقشة أكثر تفصيلاً لبعض بدائل معامل «ألفا»، مثل معامل «أوميغا»، والذي اكتسب قدراً أكبر من الاعتراف وقد حاولت أن أحافظ على أن تكون هذه المناقشة في المقام الأول مفاهيمية ونظرية وليست رياضية، وبدلاً من تكريس أو تخصيص الصفحات للعمليات الحسابية التفصيلية، فقد وجَّهت القراء إلى المصادر التي ستساعدهم في استخدام أدوات مثل معامل «أوميغا» إذا اختاروا ذلك لقد قمت أيضاً بإضافة مناقشة حول مدى ثبات درجات التغيير المحسوبة كفرق بين درجتين من مقياس واحد تم تطبيقه أكثر من مرة وألخص من الناحية النظرية ماهية ثبات درجة التغيير، وأسلط الضوء على الظروف عندما يكون في استخدامها مشكلة، وأقترح أساليب تحليلية توفر بعض الحماية ضد الثبات المخفف الناشئ عن استخدام درجات التغيير الأولية أو الخام

في الفصل الرابع، الصدق، قمت بتوسيع نطاق معالجاتي لصدق المحتوى بشكل كبير وأؤكد على أهمية الحفاظ على علاقة وثيقة بين المتغير محل الاهتمام والمقياس المستخدم لقياسه كميّاً كما أناقش أيضاً المخاطر المحتملة لوضع مفهوم للمتغير على نطاق أوسع من اللازم

وفي الفصل الخامس، القواعد الإرشادية لتطوير المقياس، قدمت عدداً قليلاً من المسائل إما مزيداً من التركيز أو قدمت مزيداً من التفاصيل إلى حد ما وحاولت أيضاً ربط المسائل العملية التي يركز عليها هذا الفصل ببعض المسائل المفاهيمية «النظرية» المثارة في فصول أخرى

في الفصل السادس، التحليل العاملي، واجهت مرة أخرى معضلة الموازنة بين الأنشطة الجديدة المهمة في هذا المجال وبين الحفاظ على الأمور لتكون في متناول القارئ وسهلة القراءة وفي النهاية، اخترت إضافة قسم جديد حول نموذج ثنائي العوامل bifactor ولقد فعلت ذلك لأنه فيما يبدو أصبح يستخدم على نطاق أوسع لتقييم المتغيرات المعقدة أو المركبة التي تشتمل على جوانب أكثر عمومية وأكثر تحديداً ولا تهدف تغطيتي إلى جعل القراء محللين مقتدرين، بل فهم سبب مواجهتهم لهذا النموذج (ونموذج العامل الهرمي المماثل من الناحية المفاهيمية) في قراءتهم وعلى وجه الخصوص في التهج القائمة على نظرية استجابة البند، يمكن للنماذج ثنائية العوامل أن تخدم غرضاً محدداً (المساعدة على تحقيق أحادية البعد) والتي شعرت أنه يستحق بعض التوضيح والتفسير

يقدم الفصل السابع مرة أخرى نظرة عامة على نظرية استجابة البند وكما فعلت مع الفصل السادس، واجهت قراراً بشأن ما يجب تضمينه ومعالجته، وكنت أرغب في إضافة المزيد من التفاصيل مع الحفاظ على الأسلوب والمسار العام للكتاب لقد اخترت تقديم مناقشة جديدة حول ثبات  $\theta$ ، أي قوة السمة التي يجري تقييمها، ودورها في اختبار التكيف المحوسب ((CAT) computerized adaptive testing ولم يصبح اختبار التكيف المحوسب CAT أكثر انتشاراً فحسب، بل إنه، في بعض النواحي، يمثل وسيلة مختصرة لتلخيص بعض الاختلافات بين نظرية استجابة البند IRT والتهج الكلاسيكية للقياس

يناقش الفصل الثامن مرة أخرى القياس في السياق الأوسع وكما فكرت في العديد من التغييرات في نهج القياس التي طرأت على أساليب القياس منذ ظهور الطبعة الأولى من هذا الكتاب، فقد أدهشتني حقيقة أن العديد منها قد جاء نتيجة لجهود القياس واسعة النطاق large-scale measurement efforts هذه الجهود جديرة بالاهتمام ومهمة ولكنها على الأرجح ليست نوع القياس الذي سيشارك فيه معظم الباحثين ولذلك، شعرت أنه سيكون من المفيد مقارنة «القياس الصغير» مع «القياس الكبير» وأؤكد في هذا الصدد على أن كليهما قادر على توفير معلومات ذات مغزى وأن لكل منهما مزاياه وعيوبه كما أؤكد على أن إبقاء هذه الاختلافات في الاعتبار منذ بداية مشروع القياس سيزيد في كثير من الأحيان من احتمال النجاح ويمكن أيضاً أن يجعل المهمة أسهل بكثير بطرق مختلفة

كما قمت بإضافة تمارين إضافية في نهاية الفصول وآمل أن تساعد هذه التمارين القراء على تقييم فهمهم للمواد، وأن تحفزهم أيضاً على مناقشة القضايا المهمة

لقد احتفظت بالميزات التي وجدها القراء مفيدة للغاية في الطبقات السابقة وهذا يعني أن التركيز يظل في المقام الأول على المفاهيم بدلاً من العمليات الحسابية وكما هو الحال في الطبقات السابقة، فإن المساحة المحدودة التي أكرسها للصيغ والمعادلات ينبغي أن تساعد القراء على رؤية ما هو أبعد من الحروف والرموز اليونانية لفهم المفاهيم التي تعبر عنها تلك الصيغ وكلما كان ذلك ممكناً، أستخدم التشبيهات لشرح العلاقات بين المفاهيم، وهو النهج الأسلوبى الذي أخبرني القراء السابقون أنه مفيد لذلك؛ ففي حين تمت إضافة المزيد من المواد «الموضوعات» المتقدمة من أجل إبقاء القراء على اطلاع على الاتجاهات الجديدة في القياس، فقد بذلت قصارى جهدي للحفاظ على إمكانية الوصول إلى الطبقات السابقة وسهولة قراءتها

شكر وتقدير:

ترغب شركة SAGE Publishing في الاعتراف بالمساهمات القيمة التي قدمها المراجعون التالية أسماؤهم: جوري كولبرت – جيرتر Jorie Colbert-Gertz، كلية جونز هوبكنز للتعليم؛ وستيفن دبليو ديتمور Stephen W Dittmore، جامعة أركنساس؛ غاري دي إيليس Gary D Ellis، جامعة تكساس إيه آند إم؛ جوستين جروس Justin H Gross، جامعة ماساتشوستس أمهيرست؛ بنيامين هيل Benjamin D Hill، جامعة جنوب ألاباما؛ وجوزيف كوش Joseph C Kush، جامعة دوكين

نبذة عن المؤلف:

قبل تقاعده في عام 2012، عمل روبرت ف ديفيليس أستاذاً في قسم السلوك الصحي (كلية جيلينجز للصحة العامة العالمية Gillings School of Global Public Health) في جامعة نورث كارولينا في تشابل هيل يتمتع الدكتور ديفيليس بخبرة تزيد عن 35 عاماً في قياس المتغيرات النفسية والاجتماعية، وقد كان عضواً نشطاً في اتحاد نظام معلومات قياس نتائج تقارير المرضى (PROMIS)، وهو عبارة عن مبادرة لخارطة طريق للمعاهد الوطنية للصحة (NIH) متعددة المواقع تهدف إلى تحديد وتعديل واختبار ونشر المقاييس النهائية لاستخدامها من قبل باحثي المعاهد

الوطنية للصحة، وكان دوره في اتحاد نظام معلومات قياس نتائج تقارير المرضى كرئيس مجال الشبكة الواسعة للنتائج والمخرجات الاجتماعية كما عمل في مجلس إدارة شعبة علم النفس الصحي التابع للجمعية الأمريكية لعلم النفس (38)، وفي قسم دراسات البحوث السريرية/النتائج/العلاجية التابع لمؤسسة التهاب المفاصل "الروماتيزم"، وفي المجلس الاستشاري لمبادرة التميز في قياس شؤون المحاربين القدامى، وقد حصل على جائزة العالم المتميز لعام 2005 من جمعية المتخصصين في أمراض الروماتيزم، وهو محرر مشارك لرعاية وبحوث التهاب المفاصل بالإضافة إلى ذلك، عمل كمحرر زائر، أو محرر مشارك زائر، أو مراجع لأكثر من عشرين مجلة أخرى وقد عمل كباحث رئيسي أو باحث مشارك منذ أوائل الثمانينيات في سلسلة من المشروعات البحثية التي تمولها الحكومة الاتحادية والمؤسسات الخاصة، وما زال يعمل بشكل نشط في دوره التحريري وكمستشار في مجموعة متنوعة من المشاريع

## الفصل الأول

### نظرة عامة

يُعد القياس من الأمور الحيوية عبر طائفة واسعة من سياقات البحوث الاجتماعية. على سبيل المثال، فُكر في الحالات الافتراضية التالية:

١- تواجه المتخصصة في علم النفس الطبي معضلة شائعة: مقياس القياس الذي تحتاجه غير موجود على ما يبدو. وتتطلب دراستها أن يكون لديها مقياس يستطيع أن يفرق ويميز بين ما يريد الأفراد حدوثه وما يتوقعون أن يحدث عندما يرون طبيباً. وتبين أبحاثها أن الدراسات السابقة استخدمت المقاييس التي أربكت وأثارت بدون قصد هاتين الفكرتين. ولا يبدو أن المقاييس الموجودة تقوم بهذا التمييز على وجه التحديد بالطريقة التي تريدها المتخصصة في علم النفس الطبي. وعلى الرغم من أنها تستطيع تكوين وإنشاء بعض الأسئلة التي يبدو إمكانية الاستفادة منها في التمييز بين ما يريده الفرد وما يتوقعه، فإنها تخشى من أن البنود أو الأسئلة «التي تم تكوينها» قد لا تكون مؤشرات موثوقة وصالحة لهذه المفاهيم.

٢- أن المتخصص في علم الأوبئة غير متأكد من كيفية المضي قدماً والمتابعة. وهو يجري تحليلات ثانوية بشأن مجموعة كبيرة من البيانات تستند إلى دراسة استقصائية وطنيه بشأن الصحة. وقال إنه يود أن يفحص العلاقة بين جوانب معينة من الإجهاد (الضغط) النفسي المدرك والحالة الصحية. على الرغم من عدم تضمين مجموعه من البنود التي تم تصميمها كمقياس للإجهاد أو الضغط النفسي في الدراسة الاستقصائية الأصلية، إلا أن العديد من البنود التي كانت تهدف في الأصل إلى قياس متغيرات أخرى يبدو أنه يمكن الاستفادة من محتواها المرتبط بالضغط النفسي. وقد يكون من الممكن تجميع هذه البنود في مقياس يتمتع بالثبات والصدق لقياس الإجهاد أو الضغط

النفسي. ومع ذلك، إذا كانت البنود المجمعة تشكل مقياساً ضعيفاً للإجهاد النفسي، فقد يتوصل الباحث إلى استنتاجات خاطئة.

٣- يشعر فريق التسويق بالإحباط في محاولاته لتخطيط حملة لخط أو مجموعة جديدة من ألعاب الأطفال الرضع عالية السعر. وقد أشارت مجموعات التركيز (focus groups) إلى أن قرارات الشراء التي يتخذها الآباء تتأثر بشده بالأهمية التعليمية الظاهرة للألعاب من هذا النوع. ويعتقد فريق التسويق أن الآباء الذين لديهم طموحات وتطلعات تعليمية ومهنية عالية لأطفالهم الرضع سيكونون أكثر اهتماماً وانجذاباً لهذا الخط الجديد من الألعاب. ولذلك، يرغب فريق التسويق في تقييم هذه التطلعات ضمن عينه كبيره ومنتشرة جغرافياً من الآباء. وتعتبر مجموعات التركيز الإضافية مرهقة للغاية للوصول إلى عينة كبيرة بما فيه الكفاية من المستهلكين.

في كل حالة من هذه الحالات، يواجه الأشخاص المهتمون ببعض المجالات العلمية مشكلة تتعلق بالقياس. ولا يهتم أي من هؤلاء الباحثين في المقام الأول بالقياس في حد ذاته. ومع ذلك، يجب على كل منهم أن يجد وسيلة لقياس ظاهره معينة قبل التصدي لتحقيق هدف البحث الرئيسي. وفي كل حالة، تكون أدوات القياس «الجاهزة أو المتوفرة» إما غير مناسبة أو غير متاحة، ويدرك جميع الباحثين أن تبني مناهج القياس العشوائية ينطوي على مخاطر إنتاج بيانات غير دقيقة. وبالتالي، فإن تطوير أدوات القياس الخاصة بهم يبدو أنه هو الخيار الوحيد المتبقي أمامهم.

وقد واجه العديد من الباحثين في العلوم الاجتماعية مشاكل مماثلة. وتتمثل إحدى الاستجابات الشائعة للغاية لهذه الأنواع من المشاكل في الاعتماد على الأدوات القائمة المشكوك في ملاءمتها. وهناك طريقة أخرى للاستجابة لهذه المشاكل تتمثل في افتراض أن بنود الاستبيان المطورة حديثاً «التي تبدو سليمة وصحيحة» ستقوم بعمل قياس مناسب. ومن الأعداء الشائعة لضعف إستراتيجيات القياس هي عدم الارتياح أو عدم الإلمام بأساليب تطوير أدوات قياس تتمتع بالثبات والصدق وعدم إمكانية الوصول إلى المعلومات العملية حول هذا الموضوع. وقد تؤدي محاولات الباحثين الرامية إلى اكتساب مهارات تطوير القياس إلى توجيههم إما إلى مصادر غامضة مخصصة أساساً للمتخصصين في القياس أو إلى المعلومات التي تكون عامة جداً بحيث لا تكون مفيدة، وتم إعداد هذا الكتاب ليكون بديلاً لتلك الخيارات.

المنظورات العامة بشأن القياس:



يُعد القياس نشاطاً أساسياً للعلوم. ونحصل على المعرفة حول الأشخاص، والأشياء، والأحداث، والعمليات من خلال ملاحظتها ورصدها، ويتطلب فهم هذه الملاحظات باستمرار قياسها بصورة كمية (بمعنى أننا نقيس الأشياء التي نهتم بها من الناحية العلمية). وتتفاعل عملية القياس والأسئلة العلمية الأوسع نطاقاً التي تخدمها مع بعضها البعض، وغالباً ما تكون الحدود بينها غير ملحوظة، ويحدث هذا، على سبيل المثال، عندما يتم اكتشاف كيان جديد أو تنقيحه في سياق القياس أو عندما يفسر المنطق الذي ينطوي عليه تحديد كيفية القياس الكمي لظاهرة موضع الاهتمام في ضوء الظاهرة نفسها. فمثلاً، قام سميث، وإيرب، وديفيليس (Smith, Earp, and DeVellis, 1995) ببحث تصورات النساء بشأن الضرب. وقد اقترح نموذج تصوري (مفاهيمي) مسبق قائم على التحليل النظري ستة مكونات متميزة لهذه التصورات. وقد أشار العمل التجريبي والذي استهدف تطوير مقياس لقياس هذه التصورات إلى أنه - بين النساء اللاتي تعرضن للضرب والنساء اللاتي لم يتعرضن له - كانت هناك تصورات أبسط بكثير هي السائدة: لقد شرح مفهوم واحد بدقة كيف استجابت النساء المشاركات في الدراسة إلى 37 بنداً من أصل 40 بنداً تم استخدامها. وتشير هذه النتيجة إلى أن ما اعتبره الباحثون مجموعة معقدة من المتغيرات كانت تنظر إليه النساء اللاتي يعشن بالمجتمع المحلي في الواقع أنه ظاهرة واحدة أوسع نطاقاً. وهكذا، فإنه في سياق استنباط وسيلة لقياس تصورات النساء حول الضرب، اكتشفنا شيئاً جديداً حول بنية تلك التصورات.

ويقول دانكن (Duncan, 1984): إن جذور القياس تكمن في العمليات الاجتماعية، وأن هذه العمليات وقياسها في الواقع تسبق العلم: "كل القياس ... هو قياس اجتماعي. والمقاييس المادية طُوّرت لأغراض اجتماعية" (ص 35). وفي إشارة إلى عمليات القياس الاجتماعي الرسمية المبكرة، مثل التصويت وإجراء التعداد، ونظم التقدم الوظيفي، ذكر «دانكن» أن أصولها تبدو وكأنها تمثل محاولات لتلبية الاحتياجات الإنسانية اليومية، وليس مجرد تجارب أجريت لتلبية الفضول العلمي. وقد ذهب إلى القول بأنه يمكن استخلاص عمليات مماثلة.

في تاريخ الفيزياء: فقد تم قياس الطول أو المسافة والمساحة والحجم والوزن والوقت من قبل الشعوب القديمة في سياق حل المشكلات العملية والاجتماعية، وقد بُنيت العلوم الفيزيائية على أسس تلك الإنجازات (ص 106).

ومهما كانت الدوافع الأولية، فإن كل مجال من مجالات العلوم يطور مجموعه إجراءات القياس الخاصة به. فقد طور علم الفيزياء، على سبيل المثال، أساليب ومعدات متخصصة للكشف عن الجزيئات شبه الذرية. وفي إطار العلوم السلوكية والاجتماعية، تطور القياس النفسي (psychometrics) كتخصص فرعي معني بقياس الظواهر النفسية والاجتماعية. وعادةً ما يكون إجراء القياس المستخدم هو الاستبانة، وتشكل متغيرات الاهتمام جزءاً من إطار نظري أوسع نطاقاً.

### الأصول التاريخية للقياس في العلوم الاجتماعية:

أمثله مبكرة:

المنطق السليم والسجل التاريخي يؤيدان ادعاء دنكان (Duncan) بأن الضرورة الاجتماعية أدت إلى تطوير القياس قبل ظهور العلم. ولا شك أن بعض أشكال القياس كانت جزءاً من ذخيرة الأنواع لدينا منذ عصور ما قبل التاريخ. فالبشر الأوائل لابد أنهم استطاعوا تقدير الأشياء والممتلكات والخصوم (المعارضين) على أساس خصائص مثل الحجم. ويذكر «دنكان» Duncan (1984) إشارات الكتاب المقدس في الاهتمام بالقياس (على سبيل المثال، «التوازن الزائف هو رجس (مكروه) عند الرب، ولكن الوزن العادل هو فرحه»، الأمثال 1:11). ويلاحظ أن كتابات أرسطو تشير إلى المسؤولين المكلفين بالتحقق من الأوزان والقياسات. وتلاحظ Anastasi (1968) أن أسلوب سقراط socratic المستخدم في اليونان القديمة ينطوي على البحث عن الفهم بطريقة يمكن اعتبارها اختباراً للمعرفة. يصف ب. هـ. ديوبوا P. H. DuBois في مقالة له عام 1964 (أعيد طبعها في بارنيت، 1976) استخدام اختبار الخدمة المدنية في وقت مبكر يصل إلى 2200 قبل الميلاد في الصين. ويستشهد رايت (1999) Wright بأمثله أخرى على الأهمية التي أولتها العصور القديمة لعملية القياس الدقيق، بما في ذلك «وزن سبعة أمثال» weight of seven الذي استندت إليها الضرائب الإسلامية في القرن السابع. كما يشير أيضاً إلى أن البعض ربط الثورة الفرنسية، جزئياً، بالفلاحين الذين سئموا من ممارسات القياس غير العادلة.

إن الفكرة القائلة بأن القياس يمكن أن ينطوي على خطأ وأنه يمكن اتخاذ خطوات معينة للحد من ذلك الخطأ تُعد رؤية أحدث. وقد أشار «بوتشوالد» Buchwald (2006) في استعراضه لتناقضات القياس وتأثيرها على المعرفة، إلى أنه على الرغم من أنه لا يزال في العشرينيات من

عمره في أواخر الستينيات من القرن السادس عشر وبداية السبعينيات من القرن السادس عشر، كان “إسحاق نيوتن” على ما يبدو هو أول من استخدم متوسط ملاحظات متعددة، وكان قصده إنتاج مقياس أكثر دقة عندما أسفرت ملاحظاته عن الظواهر الفلكية عن قيم متناقضة. ومن المثير للاهتمام أنه لم يوثق استخدام المتوسطات في تقاريره الأولية ولكنه أخفى اعتماده عليها لعقود. وقد يكون هذا الإخفاء ناجماً عن نقص وانعدام النزاهة أقل مما نشأ من مجرد الفهم المحدود للخطأ ودوره في القياس. وتعليقاً على استخفاف مماثل لعالم فلك آخر بالنسبة للملاحظات المتناقضة، يقول «ألدن» (Alder, 2002): إنه حتى في أواخر القرن الثامن عشر، لم يكن إخفاء التناقض في الملاحظة “شائعاً فقط، بل كان يُعتبر حقاً لصاحب العلم أو الموهبة. وكان يُنظر إلى الخطأ على أنه فشل أخلاقي” (ص 301). وقد خلص «بوتشوالد» (Buchwald, 2006) إلى ملاحظة مماثلة:

[طريقة عمل العلماء في القرن السابع عشر وأوائل القرن الثامن عشر] لا تعتبر الاختلافات نواتج ثانوية حتمية لعملية القياس نفسها، بل دليلاً على مهارة فاشلة أو غير كافية. ومن المحتمل أن يكون الخطأ في القياس مختلفاً قليلاً عن السلوك الخاطئ من أي نوع؛ إذ يمكن أن يترتب عليه عواقب أخلاقية، وينبغي إدارته بطرق مناسبة (صفحة 566).

لم يكن علماء الفلك الوحيدون الذين يقومون بعمليات رصد منتظمة للظواهر الطبيعية في أواخر القرن السابع عشر وأوائل القرن الثامن عشر. ففي الستينيات من القرن السادس عشر، كان “جون جراونت” (John Graunt) يقوم بتجميع معدلات الولادة والوفاة من سجلات التعميد والدفن في مدينة هامبشاير (Hampshire) بإنجلترا. واستخدم “جراونت” أسلوب حساب المتوسط (على الرغم من أنه ليس الاستخدام الشائع في هذا الوقت) لتلخيص النتائج التي توصل إليها. ووفقاً لبوتشوالد (2006) Buchwald، كان دافع Graunt لاستخدام المتوسط هو الحصول على قيمة “حقيقية” آنية. وكانت الفكرة هي أن نسبة المواليد إلى الوفيات تمثل لبعض قوانين الطبيعة، غير أن الأحداث التي لا يمكن التنبؤ بها والتي قد تحدث في أي سنة معينة سوف تخفي تلك الحقيقة الأساسية. إن وجهة النظر هذه للملاحظة كنافذة غير مكتملة في حقائق الطبيعة تشير إلى تطور متزايد في كيفية النظر إلى القياس: بالإضافة إلى القيود التي يفرضها الملاحظ، فإنه يمكن أيضاً لعوامل أخرى أن تفسد المعلومات التي يتم جمعها تجريبياً، وقد تكشف بعض التعديلات التي أدخلت على هذه القيم على نحو أكثر دقة عن الطبيعة الحقيقية للظاهرة محل الاهتمام.

وعلى الرغم من هذه الرؤى المبكرة، فإنه بعد قرن من الزمن من استخدام “نيوتن” لأول مرة لأسلوب المتوسط قبل أن يدرك العلماء على نطاق أوسع أن جميع القياسات كانت عرضة للخطأ وأن أسلوب المتوسط سوف يقلل من مثل هذا الخطأ (Buchwald, 2006). وفقاً للفيزيائي والمؤلف “ليونارد مولودينو” (Leonard Mlodinow, 2008)، في أواخر القرن الثامن عشر وأوائل القرن التاسع عشر، أجبرت التطورات في علم الفلك والفيزياء العلماء على الاقتراب من الخطأ العشوائي بشكل أكثر منهجية، مما أدى إلى ظهور الإحصاء الرياضي (mathematical statistics). وبحلول عام 1777، قارن دانييل برنولي «Bernoulli Daniel» (ابن شقيق جاكوب برنولي «Jakob Bernoulli» الأكثر شهرة) توزيعات القيم التي تم الحصول عليها من الملاحظات الفلكية على مسار سهام الرمي (s arrows' archer)، وتكتلت أو تجمعت حول نقطة مركزية تقل بشكل تدريجي على مسافات أكبر من ذلك المركز. وعلى الرغم من أن المعالجة النظرية التي رافقت تلك الملاحظة كانت خاطئة في بعض النواحي، إلا أنها تمثل بداية تحليل رسمي للخطأ في القياس (Mlodinow, 2008). ويجادل “بوتشوالد” (Buchwald, 2006) بأن أحد العيوب الأساسية في تفسيرات القرن الثامن عشر للخطأ في القياس هو الفشل في التمييز بين الخطأ العشوائي والخطأ المنتظم. وحتى بزوغ فجر القرن التالي سوف يظهر فهم أكثر دقة للعشوائية. ومع هذا الفهم المتزايد للعشوائية جاء التقدم في القياس؛ ومع تقدم القياس، تقدم العلم أيضاً.

#### ظهور الأساليب الإحصائية ودور الاختبارات العقلية:

يدعم منظور «نونالي» (Nunnally, 1978) الرأي القائل بأن الفهم الأكثر تطوراً للعشوائية والاحتمالات والإحصاء كان ضرورياً لازدهار القياس، ويقول إنه على الرغم من أن الملاحظات المنتظمة ربما كانت مستمرة، فإن عدم وجود طرق إحصائية رسمية بدرجة أكبر قد أعاق تطوير علم لقياس القدرات البشرية حتى النصف الأخير من القرن التاسع عشر. وقد بدأ تطوير الأساليب الإحصائية المناسبة في القرن التاسع عشر في النهاية من خلال مشروع “داروين” Darwin بشأن التطور وملاحظته وقياس التباين المنتظم عبر الأنواع. وقام ابن عم داروين، السير “فرانسيس غالتون” Francis Galton، بتوسيع الملاحظة المنسقة للاختلافات إلى البشر. وكان اهتمام “غالتون” الرئيسي هو وراثة الصفات التشريحية والفكرية. وكان “كارل بيرسون” Karl Pearson، الذي يعتبره الكثيرون “مؤسس الإحصاء” (على سبيل المثال، ألين وين، Allen &

3 p. 1979, Yen)، زميلاً صغيراً لجالتون. وطور بيرسون الأدوات الرياضية - بما في ذلك معامل ارتباط حاصل ضرب العزوم (Moment Correlation Coefficient-Product) والذي يحمل اسمه، أي معامل ارتباط بيرسون - اللازمة لفحص ودراسة العلاقات بين المتغيرات بشكل منتظم. يمكن للعلماء بعد ذلك أن يحددوا مدى الترابط بين الخصائص القابلة للقياس. واستمر "تشارلز سبيرمان" Charles Spearman في تقليد أسلافه ومهد الطريق للتطوير اللاحق وتعميم التحليل العاملي لاحقاً في أوائل القرن العشرين. ومن الجدير بالذكر أن العديد من المساهمين الأوائل في القياس الرسمي (بما في ذلك ألفريد بينيه Alfred Binet، الذي طور اختبارات القدرة العقلية في فرنسا في أوائل القرن العشرين) قد تقاسموا الاهتمام بالقدرات الفكرية. وبالتالي، تم تطبيق الكثير من العمل المبكر في القياس النفسي على "الاختبار العقلي".

#### دور علم النفس الفسيولوجي:

نشأ رافدٌ تاريخي آخر للقياسات النفسية الحديثة من علم النفس الفسيولوجي. وكما رأينا، كانت مشاكل القياس شائعة في علم الفلك والعلوم الطبيعية الأخرى وكانت مصدراً لاهتمام العالم إسحاق نيوتن (بوتشوالد، 2006, Buchwald). يقع علم النفس الفسيولوجي عند التقاء علم النفس والفيزياء ويهتم بالروابط والصلات بين الخصائص الفيزيائية للمحفزات أو المثيرات وكيف يدركها البشر وينظرون إليها. إن محاولات تطبيق إجراءات القياس المتبعة في الفيزياء على دراسة المشاعر (الأحاسيس) أدت إلى نقاش طويل حول طبيعة القياس. وقد لخصت "نارينس ولوسي" (Narens and Luce, 1986) هذه القضايا والمسائل. فقد لاحظ الباحثان أنه في أواخر القرن التاسع عشر، لاحظ هيلمهولتز (Helmholtz) أن الخصائص الفيزيائية (الطبيعية)، مثل الطول والكتلة، تمتلك نفس البنية الرياضية (mathematical) الحقيقية مثلها مثل الأعداد الحقيقية الموجبة<sup>1</sup>. على سبيل المثال، فإن وحدات الطول والكتلة يمكن ترتيبها وإضافة إليها مثل الأعداد العادية. وفي أوائل القرن العشرين، استمر النقاش، واعتبرت الجمعية البريطانية لتقدم العلوم (للنهوض بالعلم) أن القياس الأساسي للمتغيرات النفسية يُعد أمراً مستحيلاً بسبب المشكلات الكامنة في ترتيب أو إضافة تصورات حسية أو إدراكية. جادل «ستيفنز» (S. S. Stevens) أن قابلية الجمع / الإضافة، التي تنطبق على الطول أو الكتلة، ليست ضرورية وأشار إلى أنه يمكن للأفراد إصدار أحكام متسقة إلى حدٍ ما على شدة (قوة) الصوت. على سبيل المثال، فالأفراد يمكنهم الحكم

على صوتٍ ما بأنه أعلى من صوتٍ آخر بمقدار الضعف أو النصف. وقال “ستيفنز” إن خاصية النسبة هذه تُمكن من إخضاع البيانات المستمدة من هذه القياسات للمعالجة الرياضية. وينسب إلى “ستيفنز” تصنيف المقاييس إلى اسمية (nominal)، وترتيبية (ordinal)، وفترية (interval)، ونسبية (ratio)، وقال إن الحكم على قوة أو شدة الصوت تتوافق مع مقياس النسبة (Duncan, 1984). وفي الوقت الذي كان فيه “ستيفنز” يقدم حججه حول مشروعية قياس القياسات السيكولوجية، كان “ل.ل. ثورستون” (L. L. Thurstone) يطور الأسس الرياضية للتحليل العاملي analysis factor (Nunnally, 1978) وقد امتدت اهتمامات “ثورستون” لكل من القدرات العقلية وعلم النفس الفسيولوجي. ووفقاً لدانكان (Duncan, 1984)، فإن “ستيفنز” قد نسب إلى “ثورستون” تطبيق الطرق والأساليب السيكولوجية في قياس المحفزات الاجتماعية. وهكذا، فإن عمله يمثل تقارباً بين الجذور التاريخية المنفصلة.

التطورات اللاحقة في القياس:

تطور المفاهيم الأساسية:

وبقدر ما كان «ستيفنز» مؤثراً، فإن تصوره للقياس ليس هو الكلمة النهائية بأي حال من الأحوال. فقد عرّف القياس بأنه «تخصيص وتعيين الأرقام للأشياء أو الأحداث وفقاً للقواعد» (Duncan, 1984). وقد اعترض “دانكان” على هذا التعريف باعتباره غير مكتمل بنفس الطريقة التي يكون بها «تشغيل البيانو عن طريق ضرب مفاتيح الآلة وفقاً لبعض الأنماط» غير مكتمل. فالقياس ليس فقط تعيين الأرقام... إلخ، بل هو أيضاً تعيين الأرقام بطريقة تتوافق مع درجات مختلفة من النوعية ... أو خاصية شيء أو حدث ما (ص 126).

وقد حددت «نارينس ولوسي» (Narens and Luce, 1986) أيضاً قيوداً على تصور “ستيفنز” الأصلي للقياس وأوضحا عدداً من التحسينات اللاحقة. ومع ذلك، يؤكد عملهم على نقطة أساسية أثارها “ستيفنز” وهي: أنه توجد نماذج للقياس بخلاف النوع الذي أقرته اللجنة (الجمعية البريطانية لتقدم العلوم)، وتؤدي هذه النماذج إلى طرق قياس قابلة للتطبيق على العلوم غير الطبيعية وكذلك على العلوم الطبيعية. والخلاصة، إن هذا العمل المتعلق بالخصائص الأساسية للمقاييس قد أرسى الشرعية العلمية لأنواع وأنماط إجراءات القياس المستخدمة في العلوم الاجتماعية.

## تطور الاختبارات العقلية:

على الرغم من أن الاختبارات العقلية (أو اختبار القدرات، كما هو معروف الآن وأكثر شيوعاً)، كانت مجالاً تقليدياً نشطاً في القياس النفسي، إلا أنها لا تمثل محوراً أساسياً لهذا الكتاب. ومع ذلك، فإنه تجدر الإشارة إليها كمصدر للمساهمات الهامة في نظرية القياس وأساليبه. وقد نُشر لأول مرة في عام 1968 منشور تاريخي حول النظريات الإحصائية لدرجات الاختبارات العقلية، قام بإعداده «فريدريك م. لورد، وميلفين ر. نوفيك» M Frederic and Melvin R. Lord Novick وأعيد إصداره مؤخراً (Novick & Lord, 2008). وقد نشأ وتطور هذا العمل من الأنشطة الفكرية الثرية لمجموعة الأبحاث السيكمترية الخاصة بخدمة الاختبارات التعليمية، والتي قام بها “لورد ونوفيك”. ويلخص هذا النص المثير للإعجاب الكثير مما كان معروفاً في مجال اختبار القدرات في ذلك الوقت وكان من بين أول الأوصاف المقننة لما أصبح يعرف باسم نظرية استجابة البند (theory response item). وقد كان المدخل الأخير مناسباً بشكل خاص لمجال أوسع نطاقاً مثل الاختبارات العقلية. والعديد من أوجه التقدم في هذا الفرع من القياس النفسي أقل شيوعاً، وربما تكون أقل سهولة في التطبيق، عندما يكون الهدف هو قياس الخصائص الأخرى بخلاف القدرات العقلية. ومع مرور الوقت، أصبحت إمكانية تطبيق هذه الطرق على سياقات القياس بخلاف تقييم القدرات أكثر وضوحاً، وسنناقشها في فصل لاحق. ومع ذلك، ستركز في المقام الأول على الأساليب “الكلاسيكية أو التقليدية” التي سيطرت إلى حد كبير على قياس الظواهر الاجتماعية والنفسية غير القدرات. وهذه الطرق هي بصفة عامة أكثر مرونة وسلاسة بالنسبة لغير المتخصصين ويمكن أن تسفر عن نتائج ممتازة.

## تقييم المرض العقلي:

كان لتطور أوصاف الأمراض العقلية تاريخ منفصل يقدم دراسة حاله مفيدة في كيفية أن عدم وجود نموذج قياس إرشادي يمكن أن يجعل التقييم عملية معقدة. وعلى مر القرون، تطورت قدرة المجتمع على التعرف على أنواع مختلفة من الأمراض العقلية عن طريق الملاحظة غير المقننة تماماً نحو الجهود المبذولة لفهم العلاقات بين الأعراض، والأسباب، والعلاجات المتوافقة مع القياسات الأكثر رسميه. لقد كانت رحله صعبة.

كانت الكتابات الرومانية واليونانية والمصرية المبكرة تساوي بين ما ندركه ونعرفه على أنه أعراض للمرض العقلي مع استحواذ الجن والشياطين أو غيرها من الظروف الخارقة للطبيعة «قوى فوق الطبيعة» (على سبيل المثال، PBS, 2002). وبحلول عام 400 قبل الميلاد، كان الطبيب اليوناني أبقرات (Hippocrates) يحاول فهم الظروف العقلية على أنها ناشئة عن العمليات الفسيولوجية التي كانت محور التركيز الرئيسي في عمله العلمي (PBS, 2002). وربما كانت جهوده من أوائل الجهود التي بذلها في التفكير في المؤشرات الصريحة للمرض العقلي من حيث أسبابه الكامنة. ومع ذلك، حتى في تلك المرحلة وما بعدها، تم وصف الأمراض العقلية فينومينولوجياً<sup>(2)</sup> (بشكل ظاهري)؛ أي إن المظاهر المرتبطة بالمرض العقلي كانت مجرد بيانات وصفية بدلاً من فهمها كنقاط نهاية في تسلسل له سبب أساسي واضح واحد أو أكثر من الأسباب الكامنة.

وقد استمرت أساليب التصنيف البسيط للمرض العقلي لأكثر من ألف عام تقريباً. وقد لخص «تارتاكوفسكي» (Tartakovsky, 2011) كيف تم تصنيف المرض العقلي لأغراض التعداد في الولايات المتحدة في وقت مبكر من منتصف القرن التاسع عشر. وفي تعداد عام 1840، كانت هناك فئة واحدة فقط هي «البلاهة / الجنون» تشير إلى وجود مشكلة عقلية. وبحلول عام 1880، توسع نظام تصنيف التعداد السكاني ليشمل الفئات التالية للمرض العقلي: الهوس، والكآبة، والهوس الأحادي، والشلل الجزئي، والخرف، والهذيان، والصرع. وهذه هي الأوصاف الأساسية لحالات أو سلوكيات غير طبيعية (مثل الحزن المستمر أو الإفراط في شرب الخمر أو ضعف العضلات أو التشنجات) بدلاً من التصنيفات المسببة للأمراض.

وفي وقت مبكر من ثمانينيات القرن التاسع عشر، بدأ الطبيب النفسي الألماني «إميل كريبيلين» (Emil Kraepelin) في التمييز بشكل أكثر منهجية بين الاضطرابات العقلية. وقد كان «كريبيلين» تلميذاً لعالم النفس الألماني «فيلهلم وندت» (Wilhelm Wundt) - والذي يُعتبر مؤسس علم النفس التجريبي - وكان «كريبيلين» طبيباً أيضاً (Eysenck, 1968)، ومن ثم قدم منظورين مختلفين لتصنيفاته للأمراض العقلية. وفي عام 1883، نشر كتاب باللغة الألمانية تحت عنوان (Kraepelin, 1883) (Compendium der Psychiatric) وهو كتاب أساسي يدعو إلى تصنيف علمي بصورة أكثر للأمراض النفسية والتمييز بين الخرف بريوكوكس (داء الجنون)



والذهان الهوسي الاكتئابي. لكن مرة أخرى، على الرغم من جهوده لاستدعاء تفسيرات لهذه الأمراض، فإن فئات التشخيص المبكرة الخاصة به هي في المقام الأول وصفاً موجزًا لأعراض واضحة تميل إلى الحدوث معاً، وليست التفسيرات المسببة المقنعة (Decker, 2007). على الرغم من أن “كريبلين” قد طور النهج العلمي لفهم المرض العقلي، إلا أن الأدوات المتاحة له كانت بدائية، وفي النهاية، كانت فئاته لتشخيص الأمراض لا تزال وصفية إلى حد كبير.

ويُقيّم “ديكر” (Decker, 2007) إرث “كريبلين” على النحو التالي: “وخلصه القول: وفقاً لمعايير البحث الحالية، فإن الأعمال العلمية لـ “كريبلين” واستدلالاته من شأنها أن تثير تساؤلات حول الأفكار والمفاهيم المسبقة وتحيز الملاحظين. ويمكن رؤية أوجه القصور العلمية في وصف “كريبلين” نفسه لأساليبه ومنهجه. وعلى الرغم من تألقه في التصنيفات القنوية الواضحة، فإن تراثه العلمي قام على أسس تجريبية هشة وضعيفة” (ص 341).

في منتصف القرن العشرين، حاول الطب النفسي الأمريكي فرض المزيد من النظام على تقييم المرض العقلي. وبحلول موعد ظهور الدليل التشخيصي والإحصائي للاضطرابات العقلية (DSM؛ الرابطة الأمريكية للطب النفسي APA، 1952)، حاولت نظم التصنيف السائدة تصنيف الأمراض العقلية على أساس كل من مظاهرها ومسبباتها على حد سواء، كما هو الحال بالنسبة للصدمات الدماغية الحادة أو إدمان الكحول. ومع ذلك، لم يتم بعد تطوير مفاهيم أكثر دقة لمسببات الظروف غير المرتبطة بسبب خارجي واضح، وغالباً ما كان يفترض أسباب نفسية ديناميكية. وكان المصطلح المطبق على مثل هذه الحالات والظروف هو ردود الفعل، التي يُفترض أنها ضغوط نفسية ذات أصول غير محددة. ومره أخرى، كانت التصنيفات في المقام الأول أوصافاً لأعراض ظاهرة. وعلى الرغم من أن نظام تصنيف الدليل التشخيصي والإحصائي للاضطرابات العقلية (DSM) يمثل تقدماً واضحاً يتجاوز النظم السابقة، إلا أنه لا يزال غير مطابق لمعايير القياس الحديثة. حتى بعد أربعة عقود منذ ذلك الوقت، عندما ظهر الإصدار الرابع من الدليل التشخيصي والإحصائي للاضطرابات العقلية DSM-IV (الجمعية الأمريكية للطب النفسي، 2000)، كان هناك قدر كبير من عدم الرضا عن نظام التصنيف. وقد لاحظ عالم النفس “بول ميهل” (Paul Meehl, 1999) أن المشكلة ليست بالضرورة في استخدام الفئات (بعضها صعب وسريع، أو تنتمي أو لا تنتمي، ربما كانت هناك فئات، كما قال) ولكن عدم وجود مبرر منطقي

واضح لتعيين الناس لهذه الفئات. وعلى حد تعبير “ميهل” (1999)، “بالنسبة لتلك الأقلية من قواعد الدليل التشخيصي والإحصائي للاضطرابات العقلية التي تشير إلى الكيانات التصنيفية الحقيقية، فإن إجراءات تحديدها ومعايير تطبيقها تقتصر إلى أساس علمي كافٍ” (ص 166).

لقد أتاحت مقدمة الإصدار الخامس من الدليل التشخيصي والإحصائي للاضطرابات العقلية (DSM-V) وظهوره في نهاية المطاف في 2013 (الجمعية الأمريكية للطب النفسي، ٢٠١٣) الفرصة لإعادة النظر في تصنيف الصحة العقلية. ويرى البعض أن الفريق الذي عمل على التنقيح والمراجعة للدليل لم يستفد استفادة كاملة من تلك الفرصة التي أُتيحت له. وكما لوحظ، فإن إحدى سمات تصنيف الصحة العقلية تاريخياً هي أنها سعت إلى التصنيف بدلاً من القياس. أي إن الهدف كان وصف الوجود أو الغياب بدلاً من تحديد درجة حالة محددة. وتشير التجربة إلى أنه، حتى في الحالات، مثل الفصام، فإن “ميهل” (1999) كان على استعداد للاعتراف بالتصنيف (أي أن تكون اضطرابات منفصلة سواء كانت موجودة أو غائبة). فهناك سلسلة متصلة من الضعف بدلاً من حالة الكل أو لا شيء. ومع ذلك لا يزال الاعتماد على التصنيف بدلاً من القياس قائماً.

في العديد من الحالات، كانت تتطوي على حدود تعسفية للعلامات والأعراض، مثل اجتياز خط وهمي من الشدة يشكل وجود حالة في حين أن عدم اجتياز هذا الخط الوهمي يعبر عن عدم وجود الحالة. أيضاً، فقد استندت التصنيفات بشكل حصري تقريباً إلى ملاحظات حول الأعراض الظاهرة بدلاً من تقييمات الدلالات الرئيسية للحالات، مثل وجود الأسباب المرضية السببية أو علامة وراثية أو حالة غير طبيعية من الكيمياء الداخلية التي قد تكون أساساً لتعيين التشخيص الجسدي. عندما بدأ العمل (خارج الرؤية العامة) على الدليل التشخيصي والإحصائي للاضطرابات العقلية الإصدار الخامس DSM-V، أعرب الكثيرون عن أملهم في أن تكون المراجعة والتنقيح أكثر جرأة من الإصدارات السابقة وأن يتم تطبيق أساليب تقييم أكثر حداثة. وفي عام 2005، بعد الإعلان عن خطط مراجعة وتنقيح الدليل التشخيصي والإحصائي للاضطرابات العقلية (والتي ستصبح الإصدار الخامس من الدليل DSM-V)، بدأ المجتمع العلمي للصحة النفسية في التعبير عن مخاوفه. على سبيل المثال، ركز عدد خاص من مجلة علم النفس المرضي (Psychology Abnormal of Journal) على أهمية وجدوى إعادة التفكير في علم الأمراض

النفسية استناداً إلى تحديد الأبعاد الأساسية، مثل الفكر المضطرب، والتأثير، والسلوك، التي تؤدي إلى مشاكل محددة في الصحة العقلية (Kreuger, Watson, & Barlow, 2005). وقد جادل كروجر وآخرون (al et Kreuger, 2005) بأن هذا المدخل أو المنهج يمكن أن يعالج اثنين من أوجه القصور التجريبية الأساسية لنظم التصنيف القائمة على الفئة هما: انتشار واسع للإصابة المرضية المشتركة (أي مجموعات الأعراض الفردية التي تناسب التشخيصات المتعددة) وعدم التجانس الشديد داخل التشخيص (أي الأفراد الذين تم تحديد نفس التشخيص لهم يتشاركون في الأعراض بصورة قليلة أو ربما لا توجد أعراض). ولقد ضغط كل من الباحثين والمنظرين وحتى الفلاسفة (على سبيل المثال، أراغونا (Aragona 2009) من أجل إعادة النظر في تشخيص المرض العقلي الذي كان أكثر انسجاماً مع العمل التجريبي مثل أساليب ومداخل القياس الحديثة. ومع ذلك، وعلى الرغم من هذه الجهود، فقد أصدرت الجمعية الأمريكية للأمراض النفسية الإصدار الخامس من الدليل التشخيصي والإحصائي للاضطرابات العقلية DSM-V في شكل احتفظ بنظام التصنيف الأساسي المستخدم في الإصدارات السابقة. وقد دفع ذلك "توماس إنسل" (Thomas Insel)، مدير المعهد الوطني للصحة العقلية (NIMH)، إلى إصدار بيان على مدونته (Insel, 2013)، يقول: إن المعهد الوطني للصحة العقلية لم يعد ينظم جهوده البحثية حول فئات الدليل التشخيصي والإحصائي للاضطرابات العقلية وأنه بذل جهداً لمدة عشر سنوات، وهو مشروع معايير مجال البحث (RDoC)، وذلك لإعادة صياغة مفهوم ووضع تصور للأمراض العقلية. وقد وصف إنسل (Insel, 2013) هذا الجهد بالقول: إن مشروع معايير مجال البحث "RDoC" هو إطار لجمع البيانات اللازمة لعلم تصنيف وتشخيص الأمراض الجديد. ولكن من المهم أن ندرك أننا لا نستطيع النجاح إذا استخدمنا فئات الدليل التشخيصي والإحصائي للاضطرابات العقلية على أساس أنه "المعيار الذهبي". وفي الشهر التالي، أصدر "إنسل" بياناً صحفياً مشتركاً مع الرئيس المنتخب للجمعية الأمريكية للطب النفسي آنذاك "جيفري أ. ليبيرمان" Jeffrey A. Lieberman. في هذا الإصدار، ذكروا ما يلي:

اليوم، يمثل الدليل التشخيصي والإحصائي للاضطرابات العقلية (DSM) التابع للجمعية الأمريكية للطب النفسي (APA)، إلى جانب التصنيف الدولي للأمراض (ICD)، أفضل المعلومات المتاحة حالياً للتشخيص السريري للاضطرابات العقلية.

ومع ذلك، فإن ما قد يكون ممكناً من الناحية الواقعية في الوقت الحاضر بالنسبة للممارسين لم يعد كافياً للباحثين. واستشرافاً للمستقبل، فإن وضع الأساس لنظام تشخيصي مستقبلي يعكس بصوره مباشره علوم المخ الحديثة، سيتطلب الانفتاح لإعادة التفكير في الفئات التقليدية. ومن الواضح بشكل متزايد أن المرض العقلي سيتم فهمه على أفضل وجه على أنه اضطرابات في بنية الدماغ ووظيفته التي تتطوي على مجالات محددة من الإدراك والعاطفة والسلوك. هذا هو التركيز الخاص أو محور مشروع معايير مجال البحث (RDoC) الخاص بالمعهد الوطني للصحة العقلية. (Lieberman & Insel, 2013) NIMH.

في أكتوبر 2015، استقال “إنسل” من منصبه في المعهد الوطني للصحة العقلية (إنسل، Insel, 2015) لقبوله منصب في قسم علوم الحياة في شركة ألفا بيت (Alphabet)، والشركة هي المظلة التي تشكلت كجزء من إعادة التنظيم الهيكلي لجوجل Google. وكان من بين العوامل التي ذكرها “إنسل” أنها أثرت على قراره هو أمله في التوصل إلى منهج أكثر تنظيماً لتصنيف الصحة العقلية. كما ذكر في مقابلة لمجلة إم.آي.تي لمراجعة التكنولوجيا (Review Technology MIT)، أن انتقاله إلى شركة “ألفا بيت”، في جزء منه، يمثل “محاولة لاكتشاف طريقه أفضل لجلب تحليلات البيانات إلى الطب النفسي. فنظام التشخيص الذي نملكه يستند كلياً على الأعراض القائمة وذات طابع شخصي إلى حد ما” (Regalado, 2015). ويأمل الكثيرون في أن العمل الذي يقوم به “إنسل” في شركة “ألفا بيت” سوف يعزز تحديث التقييم النفسي لجعله أكثر توافقاً مع معايير القياس الحديثة.

#### توسيع نطاق القياس النفسي:

يذكر دنكان (Duncan, 1984) أن تأثير القياس النفسي في العلوم الاجتماعية قد تجاوز أصوله في قياس الأحاسيس والقدرات الفكرية. ومن الواضح أن القياس النفسي ظهر كنموذج منهجي أو ميثودولوجي في حد ذاته. ويدعم “دنكان” هذه الحجة بثلاثة أمثلة لتأثير القياسات النفسية: (1) الاستخدام واسع الانتشار للتعريف السيكمترية للثبات والصدق، (2) انتشار استخدام التحليل العاملي في أبحاث العلوم الاجتماعية، و(3) تبني أساليب القياس النفسي من أجل تطوير المقاييس التي تقيس مجموعة من المتغيرات أوسع بكثير من تلك التي كانت لا تهتم بها القياسات

النفسية في البداية (ص 203). إن قابلية تطبيق المفاهيم والأساليب السيكمترية لقياس الظواهر النفسية والاجتماعية المتنوعة ستستحوذ على اهتمامنا فيما تبقى من هذا الكتاب.

دور القياس في العلوم الاجتماعية:

علاقة النظرية بالقياس:

الظواهر التي نحاول قياسها في أبحاث العلوم الاجتماعية غالباً ما تتبع من النظرية. وبالتالي، تلعب النظرية دوراً رئيسياً في كيفية تصورنا لمشاكل القياس الخاصة بنا. في الواقع، يذكر لورد ونوفيك (Lord and Novick, 2008) أن للقضايا النظرية دوراً مهماً في تطوير نظرية القياس. وكان المنظرون يشعرون بالقلق من أن تقديرات العلاقات بين المفاهيم البنائية constructs محلاً للاهتمام، قد تم الحصول عليها عمومًا من خلال المؤشرات المرتبطة بتلك المفاهيم البنائية. ونظراً لأن تلك المؤشرات تحتوي على خطأ، فإن الارتباطات الناتجة عن ذلك كانت أقل من العلاقة الفعلية بين المفاهيم البنائية، وقد حفز ذلك على تطوير واستحداث أساليب لتعديل وتصحيح الارتباطات من أجل تقليل الخطأ وحفز تطوير نظرية القياس كمجال اهتمام وتركيز متميز (ص 69).

بطبيعة الحال، فإن العديد من مجالات العلوم تقيس الأشياء المشتقة أو المستمدة من النظرية. وحتى يتم تأكيد العناصر الفرعية من خلال القياس، فهي أيضاً مجرد مفاهيم بنائية نظرية. ومع ذلك، تختلف النظرية في علم النفس والعلوم الاجتماعية الأخرى عن النظرية في العلوم الطبيعية. ويميل العلماء في العلوم الاجتماعية إلى الاعتماد على العديد من النماذج النظرية التي تتعلق أو تهتم بالظواهر الضيقة النطاق والمحددة إلى حد ما، في حين أن النظريات في العلوم الطبيعية أقل من حيث العدد وأكثر شمولاً من حيث النطاق أو المجال. على سبيل المثال، تركز نظرية المقارنة الاجتماعية لفستنجر (Festinger, 1954) على مجموعة ضيقة إلى حد ما من الخبرة الإنسانية: الطريقة التي يُقِيم بها الناس قدراتهم أو آراءهم من خلال مقارنة أنفسهم مع الآخرين. وعلى النقيض من ذلك، يواصل علماء الفيزياء العمل نحو إيجاد نظرية موحدة كبيرة النطاق تضم جميع القوى الأساسية للطبيعة ضمن إطار مفاهيمي واحد. أيضاً، العلوم الاجتماعية أقل نضجاً من العلوم الطبيعية، ونظرياتها التي تتطور بسرعة أكبر. ويشكل قياس الظواهر المعقدة وغير الملموسة المستمدة من نظريات متعددة ومتطورة، تحدياً واضحاً للباحثين في العلوم الاجتماعية. ولذلك، من

المهم بصفة خاصة أن نضع في اعتبارنا إجراءات القياس وأن ندرك تماماً نقاط قوتها وأوجه القصور فيها.

وكما ازدادت معارف الباحثين عن الظواهر التي يهتمون بدراستها، والعلاقات التجريدية الموجودة بين المفاهيم البنائية الافتراضية، والأدوات الكمية المتاحة لهم، كانوا أكثر جاهزية بشكل أفضل لتطوير مقاييس تتميز بالثبات والصدق والقابلية للاستخدام والتطبيق. ولعل المعرفة التفصيلية بظاهرة محددة موضع الاهتمام تُعد أهم هذه الاعتبارات. على سبيل المثال، نظرية المقارنة الاجتماعية لها العديد من الجوانب التي قد تنطوي على إستراتيجيات قياس مختلفة. وقد يتطلب أحد الأسئلة البحثية تفعيل المقارنات الاجتماعية باعتبارها تفضيلاً نسبياً للمعلومات المتعلقة بالآخرين ذوي المكانة الأعلى أو الأدنى، في حين قد يفرض سؤال بحثي آخر تصنيف الذات بالنسبة إلى "الشخص العادي" على مختلف الأبعاد. وقد لا تسفر المقاييس المختلفة التي تقيس جوانب مختلفة من نفس الظاهرة العامة (مثل المقارنة الاجتماعية) عن نتائج متقاربة (ديفيليس، al et DeVellis, 1990). وفي الحقيقة، تقوم المقاييس بتقييم متغيرات مختلفة على الرغم من استخدام اسم متغير مشترك أو عام في توصيفها. وبالتالي، فإن وضع وتطوير مقياس مناسب على النحو الأمثل لسؤال البحث يتطلب فهم التفاصيل الدقيقة للنظرية.

تتطلب المتغيرات المختلفة إستراتيجيات تقييم مختلفة. على سبيل المثال، يمكن ملاحظة عدد العملات المميزة المأخوذة مباشرة من صندوق أو حاوية، في حين أنه لا يمكن ملاحظة الكثير من المتغيرات التي يهتم بها العلماء الاجتماعيون والسلوكيون بشكل مباشر؛ فالمعتقدات، والحالات التحفيزية، والتوقعات، والاحتياجات، والعواطف، وتصورات الدور الاجتماعي ليست سوى أمثلة قليلة. لا يمكن ملاحظة بعض المتغيرات مباشرة ولكن يمكن تحديدها بإجراءات بحثية غير الاستبيانات. على سبيل المثال، على الرغم من أنه لا يمكن للباحثين المعرفيين أن يلاحظوا بشكل مباشر كيفية تنظيم الأفراد للمعلومات حول النوع الاجتماعي في مخططاتهم الذاتية، إلا أنهم قد يكونون قادرين على استخدام إجراءات التذكر للاستدلال على كيفية تنظيم الأفراد لأفكارهم حول الذات والجنس. ومع ذلك، هناك العديد من الحالات التي يستحيل فيها أو لا يكون عملياً تقييم متغيرات العلوم الاجتماعية بأي طريقة أخرى غير استخدام مقياس القياس الورقي والقلم الرصاص. هذا هو الحال في كثير من الأحيان، ولكن ليس دائماً، عندما يكون الاهتمام منصباً على قياس

المفاهيم البنائية. وهكذا، فإن الباحث المهتم بقياس ظاهرة أنثوية المظهر androgyny قد يجد أنه من الأسهل بكثير القيام بذلك عن طريق استبانة تم تطويرها بعناية من خلال إجراءات بديلة.

#### المقاييس النظرية وغير النظرية:

في هذه المرحلة، ينبغي أن نعترف بأنه على الرغم من أن هذا الكتاب يركز على مقاييس المفاهيم البنائية، فليس من الضروري أن تكون كل التقييمات باستخدام الورقة والقلم الرصاص في حاجة إلى نظرية. فالجنس والسن، على سبيل المثال، يمكن التحقق منهما من التقرير الذاتي عن طريق استبانة. وبناءً على سؤال البحث، يمكن أن يكون هذان المتغيران من مكونات نموذج نظري أو ببساطة جزءاً من وصف المشاركين في الدراسة. لا يوجد أساس نظري لبعض السياقات التي يُطلب فيها من الأشخاص الإجابة على قائمة من الأسئلة باستخدام نسق أو شكل ورقي وقلم رصاص، مثل تقييم تفضيلات وجبة المريض في المستشفى. وفي حالات أخرى، قد تبدأ الدراسة غير نظرية ولكنها تؤدي إلى صياغة النظرية. على سبيل المثال، قد يطلب باحث في مجال التسوق من الآباء سرد أنواع الألعاب التي اشتروها لأطفالهم. بعد ذلك، قد يستكشف الباحث من هذه القوائم أنماط العلاقات. وبناءً على الأنماط الملاحظة لعمليات شراء الألعاب، يمكن للباحث تطوير نموذج لسلوك الشراء. وتمثل استبانات الرأي العام مثلاً آخر على القياس اللانظري إلى حدٍ ما. إن سؤال الناس عن العلامة التجارية للصابون الذي يستخدمونه أو لمن يعتزمون التصويت لصالحهم في الانتخابات نادراً ما ينطوي على أي محاولة للاستفادة من مفهوم بنائي نظري أساسي. وبدلاً من ذلك، فإن الاهتمام يكمن في الاستجابة الموضوعية في حد ذاتها، وليس في بعض خصائص الشخص التي يفترض أن يعبر عنها.

قد يكون التمييز بين مواقف القياس النظرية وغير النظرية أمراً صعباً في بعض الأحيان. على سبيل المثال، فإن البحث عن تفضيل الناخب في المرشحين للرئاسة كوسيلة للتنبؤ بنتيجة الانتخابات يعني مطالبة المجيب أو المجيبة بالإبلاغ عن نيته أو نيتها السلوكية. قد يسأل الباحث الناس عن كيفية تخطيطهم للتصويت في الانتخابات، ليس بدافع الاهتمام بعمليات صنع القرار الانتخابي بل لمجرد توقع نتائج الانتخابات النهائية. ومن ناحية أخرى، إذا طرح السؤال نفسه في سياق دراسة حول كيفية تأثير الاتجاهات أو المواقف نحو قضايا محددة على تفضيل المرشحين، فإن وجود نظرية مفصلة بشكل جيد قد تكون أساساً للبحث. ولا تهدف المعلومات المتعلقة

بالتصويت في هذه الحالة إلى الكشف عن الكيفية التي سوف يصوت بها المستجيبون، ولكن لتسليط الضوء على الخصائص الفردية. في هاتين الحالتين، فإن صلة أو عدم صلة المقياس بالنظرية تُعد مسألة تتعلق بنية وهدف الباحث، وليست بالإجراءات المستخدمة. وتتم إحالة القراء المهتمين بمعرفة المزيد حول بناء استبانات المسح التي ليست معنية في المقام الأول بقياس المفاهيم البنائية الافتراضية إلى المراجع التالية: كونفيرس، وبيريسر (1986) Converse and Presser؛ كزاجا وبلير Blair and Czaja (1996)؛ ديلمان (2007) Dillman؛ فينك (1995) Fink؛ فاوولر Fowler (2006)؛ وفيسبرج، وكروسنيك، وبراون Weisberg, Krosnick, and Bowen (1996).

#### أنواع المقاييس:

غالباً ما يشار إلى أدوات القياس التي تتكون من مجموعات من الفقرات، والتي يتم تجميعها معاً في درجه مركبة بهدف الكشف عن مستويات المتغيرات النظرية التي لا يمكن ملاحظتها بسهولة بالوسائل المباشرة، على أنها مقاييس. ونحن نطور مقاييس عندما نريد قياس الظواهر التي نعتقد أنها موجودة بسبب فهمنا النظري للعالم، ولكننا لا نستطيع أن نقيّمها بصورة مباشرة. على سبيل المثال، قد نقوم باستحضار الاكتئاب أو القلق كتفسيرات للسلوكيات التي نلاحظها. ويتفق معظم المنظرين على أن الاكتئاب أو القلق لا يعادل السلوك الذي نراه ولكنه يكمن وراءه. وتفترض نظرياتنا أن هذه الظواهر موجودة وأنها تؤثر على السلوك ولكنها غير ملموسة. وفي بعض الأحيان، قد يكون من المناسب استنتاج وجودها من عواقبها السلوكية. ومع ذلك، في بعض الأحيان، قد لا نتمكن من الوصول إلى المعلومات السلوكية (كما هو الحال عندما تقتصر على استخدام منهجيات المسح البريدي)، أو قد لا نكون متأكدين من كيفية تفسير عينات أو نماذج السلوك المتاحة (كما هو الحال عندما يظل الشخص سلبياً في مواجهة حدث قد يتفاعل معه معظم الأشخاص الآخرين بقوة)، أو قد لا يرغب في افتراض أن السلوك غير متماثل مع المفهوم البنائي الأساسي محل الاهتمام (كما هو الحال عندما نشك في أن البكاء هو نتيجة الفرح وليس الحزن). وفي الحالات التي لا نستطيع فيها الاعتماد على السلوك كمؤشر على وجود ظاهرة ما، قد يكون من المفيد تقييم المفهوم البنائي من خلال مقياس تم بناؤه بعناية والتحقق من صدقه.



وحتى بين المتغيرات المشتقة من النظرية، هناك سلسلة متصلة ضمنية تتراوح من ظواهر ملموسة نسبياً يمكن الوصول إليها إلى الظواهر المجردة نسبياً والتي يتعذر الوصول إليها. ولن يتطلب كل شيء مقاييس متعددة البنود أو الفقرات. ومن المؤكد أن العمر والجنس لهما صلة بالعديد من النظريات ولكن نادراً ما يتطلبان مقياساً متعدد البنود لإجراء تقييم دقيق. يعرف الناس سنهم وجنسهم. وترتبط هذه المتغيرات، في معظم الأحوال، بخصائص ملموسة لا لبس فيها نسبياً (مثل التشريح) أو الأحداث (مثل تاريخ الميلاد). وما لم تكن هناك ظروف خاصة مثل ضعف الجهاز العصبي، يمكن للمستجيبين استرجاع وتذكر المعلومات المتعلقة بأعمارهم وجنسهم من الذاكرة بسهولة تامة. ويمكنهم الاستجابة بدرجة عالية من الدقة على سؤال واحد لتقييم متغيرات مثل هذه. ويمكن القول إن الأصل العرقي أكثر تعقيداً وتجريداً من العمر أو الجنس، وعادةً ما ينطوي على مزيج من العوامل المادية والثقافية والتاريخية. ونتيجة لذلك، فإن المفهوم البنائي الاجتماعي أقل في درجة الملموسية - من العمر أو الجنس. وعلى الرغم من أن الآليات التي ينطوي عليها تحديد الشخص الأصل العرقي قد تكون معقدة وتتكشف على مدى فترة زمنية طويلة، فإن معظم الأفراد توصلوا إلى تعريف شخصي ويمكنهم الإبلاغ عن أصلهم العرقي بقليل من التفكير أو التأمل. وبالتالي، قد يكفي متغير واحد لتقييم الأصل العرقي في معظم الظروف والحالات (غير أن هذا قد يتغير لأن مجتمعنا أصبح تدريجياً أكثر تعدداً في العرقيات ولأن الأفراد يعرفون انتماءهم العرقي الشخصي من حيث المجموعات الاثنية المتعددة التي تعكس أسلافهم). ومع ذلك، فإن العديد من المتغيرات النظرية الأخرى تتطلب من المستجيب إعادة بناء المعلومات القليلة المتاحة أو تفسيرها أو الحكم عليها أو مقارنتها أو تقييمها. على سبيل المثال، قياس كيف يعتقد الأشخاص المتزوجون أن حياتهم ستكون مختلفة إذا اختاروا زوجاً مختلفاً، على الأرجح يتطلب جهداً عقلياً وذهنياً كبيراً، وقد لا يستطيع بند واحد الكشف عن تعقيد الظاهرة محل الاهتمام. وفي ظروف كهذه، قد يكون المقياس هو أداه التقييم الأكثر ملاءمة. وقد تتمكن بنود متعددة من الكشف عن جوهر مثل هذا المتغير بدرجة من الدقة التي لا يمكن أن يحققها بند واحد. وهذا النوع من المتغيرات تحديداً - وهو المتغير الذي لا يمكن ملاحظته مباشرة والذي ينطوي على التفكير من جانب المستجيبين - هو الذي يتم تقييمه بشكل مناسب من خلال مقياس.

وينبغي مقارنة المقياس بأنواع أخرى من المقاييس المتعددة البنود التي تسفر عن درجة مركبة. والتميز بين هذه الأنواع المختلفة من البنود المركبة ذات أهميه نظرية وعملية على حد

سواء، كما ستكشف الفصول اللاحقة. ووفقاً للمصطلحات المستخدمة في هذا الكتاب، فإن المقياس يتكون كما أشار «بولين» (Loehlin, 1998, pp. 200-202؛ انظر أيضاً 65 - pp. 64 1989) من «مؤشرات التأثير» effect indicators - وهي البنود التي تكون قيمتها ناتجة من المفهوم البنائي الأساسي (أو متغير كامن، كما سنشير إليها في الفصل التالي). وغالباً ما يتفق مقياس الاكتئاب مع خصائص المقياس، حيث إن الاستجابات على كل بند من البنود على حدة التي تتقاسم أو تشترك في قضية عامة - وتسمى الحالة الوجدانية للمستجيب. وهكذا، فإن كيفية استجابة شخص ما لبنود مثل «أشعر بالحزن» و«حياتي كئيبة» ربما تحددها إلى حد كبير مشاعر ذلك الشخص في ذلك الوقت. سأستخدم مصطلح «دليل» index، من ناحية أخرى، لوصف مجموعات من البنود التي تسبب مؤشرات - أي البنود التي تحدد مستوي مفهوم بنائي ما. فعلي سبيل المثال، من الممكن أن يكون قياس جاذبية المرشح الرئاسي مناسباً لخصائص الدليل index. وقد تقوم هذه البنود بتقييم مكان الإقامة الجغرافية للمرشح، وحجم أسرته، وجاذبيته المادية، وقدرته على إلهام العاملين في الحملة الانتخابية، والموارد المالية المحتملة. وعلى الرغم من أن هذه الخصائص ربما لا تشترك في أي قضية عامة، إلا أنها قد تتشارك جميعها في التأثير - مما يزيد من احتمال نجاح الحملة الرئاسية. فالبنود ليست نتيجة لشيء واحد، ولكنها تحدد النتيجة نفسها. وهناك مصطلح أكثر عمومية لمجموعه من البنود التي يمكن تجميعها في درجة مركبة هو المتغير الناشئ emergent variable (على سبيل المثال، Cohen, Cohen, Teresi, Marchi, & Velez, 1990)، والذي يتضمن مجموعات من الكيانات التي تشترك في خصائص معينة ويمكن تجميعها تحت عنوان فئة مشتركة. غير أن تجميعهم معاً لا يعني بالضرورة وجود أي علاقة سببية. فالجمل التي تبدأ بكلمة لها أقل من خمسة أحرف، على سبيل المثال، يمكن تصنيفها بسهولة معاً على الرغم من أنها لا تشترك في قضية مشتركة ولا في أي أثر مشترك. إن المتغير الناشئ ينبثق لمجرد أن شخصاً ما أو شيئاً ما (مثل برنامج تحليل البيانات) يرى نوعاً من التشابه بين البنود المعنية.

#### جميع المقاييس ليست متساوية:

للأسف، لم يتم تطوير جميع البنود المركبة بعناية. وبالنسبة للكثيرين، قد يكون التجميع مصطلح أكثر ملاءمة من التطوير. غالباً ما يقوم الباحثون بحذف البنود أو تجميعها معاً وافترض أنها تشكل مقياساً مناسباً. قد لا يفكر هؤلاء الباحثون فيما إذا كانت البنود تشترك في قضية عامة

(وبالتالي تشكل مقياساً scale)، أو تشترك في إحداث نتيجة أو عاقبة مشتركة (وبالتالي تشكل دليل index)، أو مجرد أمثلة لفئة أهداف مرتبة مشتركة لا تنطوي على أي مقدمات سببية عامة أو نتيجة (وبالتالي تشكل متغيراً ناشئاً).

لا يفشل الباحث في استغلال النظرية في تطوير المقياس فحسب، بل قد يتوصل أيضاً إلى استنتاجات خاطئة حول النظرية من خلال إساءة تفسير ما يقيسه المقياس. وهناك أمر مؤسف ولكنه متكرر بشكل مقلق للغاية هو استنتاج أن بعض المفاهيم البنائية غير مهمة أو أن نظرية معينة غير متناسقة بناءً على أداء مقياس قد لا يعكس المتغير الذي يفترضه الباحث. لماذا قد يحدث هذا؟ في البحوث نادراً ما نقوم بدراسة العلاقات بين المتغيرات بشكل مباشر. فكما ذكرنا سابقاً، فإن العديد من المتغيرات محل الاهتمام لا يمكن ملاحظتها مباشرة، وهذه حقيقة يمكننا نسيانها بسهولة. وفي كثير من الأحيان، فإننا نقوم بتقييم العلاقات بين ممثلين عن المتغيرات أو «وكلاء» proxies (مثل المقاييس) التي تهدف إلى تمثيل متغيرات محل الاهتمام. قد يصبح هناك خلط بين الوكيل أو الممثل proxy القابل للملاحظة والمتغير الذي لا يمكن ملاحظته. على سبيل المثال، يبدو أن متغيرات مثل ضغط الدم ودرجة حرارة الجسم، في الاعتبار الأول، يمكن ملاحظتهما بشكل مباشر، ولكن ما نلاحظه فعلياً عبارة عن وكلاء، مثل عمود الزئبق. تفترض استنتاجاتنا حول المتغيرات أن الوكلاء الذين تتم ملاحظتهم يرتبطون ارتباطاً وثيقاً بالمتغيرات الأساسية التي يمثلونها. هذا هو الحال بالنسبة لمقياس الحرارة؛ إننا نصف مستوى الزئبق في مقياس الحرارة بأنه «درجة الحرارة»، على الرغم من أنه بالمعنى الدقيق للكلمة، هو مجرد مظهر مرئي لدرجة الحرارة (أي الطاقة الحرارية). في هذه الحالة، حيث يتطابق الطرفان عن كثب، تكون عواقب الإشارة إلى القياس (قيمة المقياس التي يصل إليها الزئبق) كمتغير (كمية الطاقة الحرارية) تكاد تكون دائماً غير ذات أهمية. عندما تكون العلاقة بين المتغير ومؤشره أضعف مما هي عليه في مثال مقياس الحرارة، فإن الخلط بين المقياس والظاهرة التي يهدف إلى كشفها يمكن أن يؤدي إلى استنتاجات خاطئة. لنأخذ في الاعتبار حالة افتراضية يرغب فيها الباحث في إجراء تحليل ثانوي لمجموعة بيانات موجودة. ولنفترض أن هذا الباحث مهتم بدراسة دور الدعم الاجتماعي في التحصيل المهني اللاحق. ويلاحظ الباحث أن مجموعة البيانات المتاحة تحتوي على ثروة من المعلومات حول الأوضاع المهنية للأشخاص على مدى فترة زمنية طويلة وأن الأشخاص قد تم سؤالهم عما إذا كانوا متزوجين أم لا. في الواقع، قد يكون هناك العديد من البنود التي تم جمعها في أوقات مختلفة، والتي تتعلق بالزواج. ولنفترض

كذلك، في حالة عدم وجود أي بيانات توفر تقييماً أكثر تفصيلاً للدعم الاجتماعي، يقرر الباحث جمع بنود الزواج هذه في “مقياس” وأن يستخدم ذلك كمقياس للدعم. ويتفق معظم علماء العلوم الاجتماعية على أن مساواة الدعم الاجتماعي بالحالة الزوجية ليس له ما يبرره. حيث يسقط هذا الأخير ويلغي جانبين مهمين من جوانب الدعم الاجتماعي (على سبيل المثال، إدراك نوعية أو جودة الدعم الذي يتم الحصول عليه) ويتضمن عوامل يحتمل أنها غير ذات صلة بالموضوع (على سبيل المثال، حالة طفل صغير جداً بحيث لا يتمكن من الزواج مقابل شخص بالغ في سن مناسب للزواج في وقت القياس). إذا استنتج هذا الباحث الافتراضي، استناداً إلى طريقة التقييم هذه، أن الدعم الاجتماعي لم يلعب أي دور في تحقيق التحصيل المهني، فإن هذا الاستنتاج قد يكون خاطئاً تماماً. في الواقع، كانت المقارنة بين الحالة الزوجية والتحصيل المهني (أو، على نحو أدق، مؤشرات هذين المتغيرين). ولا يكون الاستنتاج المتعلق بالدعم والتحصيل المهني صحيحاً وصادقاً إلا إذا أشار الزواج فعلاً إلى مستوى الدعم.

#### تكلفة سوء أو ضعف القياس:

وحتى إذا كان القياس السيئ هو المقياس الوحيد المتاح، فإن تكاليف استخدامه قد تكون أكبر من أي فوائد يمكن تحقيقها. والحالات نادرة في العلوم الاجتماعية التي يجب أن يُتخذ فيها قرار فوري من أجل تجنب العواقب الوخيمة، وليس أمام المرء خيار آخر سوى القيام باستخدام أفضل الأدوات المتاحة. ولكن حتى في هذه الحالات النادرة، لا تختفي وتتلاشى المشاكل المتأصلة في استخدام المقاييس الرديئة لتقييم المفاهيم البنائية. إن استخدام قياس لا يقوم بتقييم ما يفترضه المرء يمكن أن يؤدي إلى قرارات خاطئة. هل هذا يعني أننا ينبغي أن نستخدم فقط أدوات القياس التي خضعت للتطوير الدقيق واختبار الصدق الشامل؟ وعلى الرغم من أن القياس المعيب أو الناقص قد يكون أفضل من عدم وجود قياس على الإطلاق في بعض الحالات، فإنه ينبغي أن ندرك متى تكون إجراءات القياس الخاصة بنا بها عيوب وأن نقلل من استنتاجنا وفقاً لذلك.

في كثير من الأحيان، سوف يعتبر الباحث القياس كمسألة ثانوية لقضايا علمية أكثر أهمية والتي تكون حافزاً على الدراسة، وبالتالي، سيحاول الباحث أن يقتصد من خلال الحد من عملية القياس. ومع ذلك، فإن المقاييس المناسبة هي شرط ضروري لإجراء بحوث سليمة وصادقة. ويجب على الباحثين السعي لتحقيق التماثل بين المفاهيم البنائية النظرية محل اهتمامهم وطرق

القياس التي يستخدمونها لقياسها إجرائياً. ويفرض سوء أو ضعف القياس قيوداً كبيرة على صدق الاستنتاجات التي يمكن الوصول إليها. بالنسبة للباحث الذي يفضل أن يولي قدراً ضئيلاً من الاهتمام للقياس وأكبر قدر ممكن للقضايا الموضوعية، فإن الإستراتيجية المناسبة قد تتمثل في تصحيح الجزء الخاص بالقياس منذ البداية المبكرة للبحث حتى يمكن اعتباره أمراً مفروغاً منه أو مسلماً به إلى حد ما بعد ذلك.

يمكن للباحث أيضاً أن يقتصد بشكل خاطئ من خلال استخدام المقاييس المختصرة جداً على أمل تخفيف العبء على المستجيبين. إن الاستبانة التي تكون مختصرة جداً لكي تتمتع بالثبات تُعد فكرة سيئة بصرف النظر عن عدد المستجيبين الذين يفضلون الإيجاز. وينتج عن الاستبانة التي تتمتع بالثبات ويتم إكمال الإجابة عليها من قبل نصف المستجيبين معلومات أكثر من الاستبانة التي لا تتمتع بالثبات وتم الحصول على إجابة جميع المستجيبين عليها. إذا لم تتمكن من تحديد ما تعنيه البيانات، فكمية المعلومات التي يتم جمعها تكون غير ذات صلة وغير مناسبة. وبالتالي، فإن استكمال الإجابة على الاستبانات «الملائمة» التي لا يمكن أن تسفر عن معلومات ذات مغزى، هو استخدام سيئ لوقت المستجيبين وجهدهم مقارنةً باستكمالهم نسخة أطول إلى حد ما تنتج بيانات صادقة. وبالتالي، فإن استخدام أساليب وطرق تقييم مختصرة وغير كافية قد يكون لها آثار أخلاقية وكذلك آثار علمية.

#### الملخص والمراجعة:

يؤكد هذا الفصل على أن القياس نشاط أساسي في جميع فروع العلوم، بما في ذلك العلوم السلوكية والاجتماعية. وللقياس النفسي، وهو مجال التخصص في العلوم الاجتماعية المعني بقياس الظواهر الاجتماعية والنفسية، سوابق تاريخية تمتد إلى العصور القديمة. في العلوم الاجتماعية، تلعب النظرية دوراً حيوياً في تطوير القياس والمقاييس، والتي هي مجموعات من البنود التي تكشف عن مستوى متغير نظري أساسي. ومع ذلك، لا تشكل جميع مجموعات البنود مقاييس بهذا المعنى. وقد يكون تطوير المقاييس أكثر صعوبة من اختيار البنود بشكل عرضي؛ ومع ذلك، فإن تكاليف استخدام المقاييس المبنية بشكل عرضي عادةً ما تفوق الفوائد والمنافع المتحققة إلى حد كبير.

تغطي الفصول التالية الأساس المنطقي وأساليب تطوير المقياس بمزيد من التفصيل. حيث يستكشف الفصل الثاني المتغير الكامن، وهو المفهوم الأساسي الذي يحاول المقياس قياسه كمياً،

ويعرض الأسس النظرية للأساليب التي تم وصفها في الفصول اللاحقة. ويقدم الفصل الثالث أساساً مفاهيمياً لفهم الثبات والمنطق الكامن وراء معامل الثبات. ويستعرض الفصل الرابع الصلاحية أو الصدق، في حين أن الفصل الخامس هو دليل عملي للخطوات التي ينطوي عليها وضع وتطوير المقياس. ويقدم الفصل السادس مفاهيم التحليل العاملي ويصف استخدامها في تطوير المقياس. والفصل السابع هو نظرة عامه مفاهيمية لمدخل بديل لتطوير المقياس - نظرية استجابة البند. وأخيراً، يناقش الفصل الثامن بإيجاز كيفية تناسب المقاييس مع عمليه البحث الأوسع نطاقاً.

تمارين:

1- ما هي الاختلافات الرئيسية بين المقياس (scale) والدليل (index) كما تم وصفهم؟

2- هناك مهنتان لهما تاريخ طويل في التقييم هما التعليم (من خلال تطوير واستخدام اختبارات القدرة الموحدة) والطب النفسي (من خلال مواصفات وتطبيق معايير التشخيص المعياري أو الموحد). ما هي بعض الاختلافات الرئيسية بين كيفية اقتراب هذين المجالين الباحثين من مجال التقييم؟

## الفصل الثاني

### فهم المتغير الكامن

يقدم هذا الفصل مخططاً وتصوراً مفاهيمياً لفهم العلاقة بين المقاييس والمفاهيم البنائية constructs التي تمثلها، على الرغم من أنه ليس الإطار الوحيد المتاح. وتُعد نظرية استجابة البند منظور قياس بديل والذي سنقوم بدراسته في الفصل السابع. ونظرًا لأنه إطار مفاهيمي مهم ويمكن حسابه واستخدامه على نطاق واسع، فإنني أؤكد على نموذج القياس الكلاسيكي، الذي يفترض أن كل بند من البنود على حدة هي مؤشرات قابلة للمقارنة مع المفاهيم البنائية الأساسية.

#### المفاهيم البنائية (constructs) مقابل المقاييس:

عادةً ما يهتم الباحثون بالمفاهيم البنائية بدلاً من البنود أو المقاييس في حد ذاتها. على سبيل المثال، سيكون باحث في مجال التسويق يهدف إلى قياس تطلعات الوالدين تجاه أطفالهم أكثر اهتماماً بالمشاعر الأبوية غير الملموسة والآمال فيما سينجزه أطفالهم مقارنةً بالمكان الذي يضع فيه هؤلاء الآباء علامات على الاستبانة. ومع ذلك، فإن تسجيل الاستجابات على استبانة قد يكون، في كثير من الحالات، أفضل طريقة لتقييم تلك المشاعر والآمال. وعادةً ما تكون بنود المقياس وسيلة نهائية لتقييم المفهوم البنائي. وبعبارة أخرى، الاستبانة ضرورية لأن العديد من المفاهيم البنائية لا يمكن تقييمها مباشرةً. بمعنى، أن المقاييس تعتبر وكلاء (proxies) للمتغيرات التي لا يمكننا ملاحظتها بشكل مباشر. ومن خلال تقييم العلاقات بين المقاييس، نستنتج بشكل غير مباشر العلاقات بين المفاهيم البنائية. في الشكل رقم (2-1)، على سبيل المثال، على الرغم من أن اهتمامنا الأساسي هو العلاقة بين المتغيرين A و B، فإننا نقدر تلك العلاقة على أساس العلاقة بين المقاييس المقابلة لتلك المتغيرات.

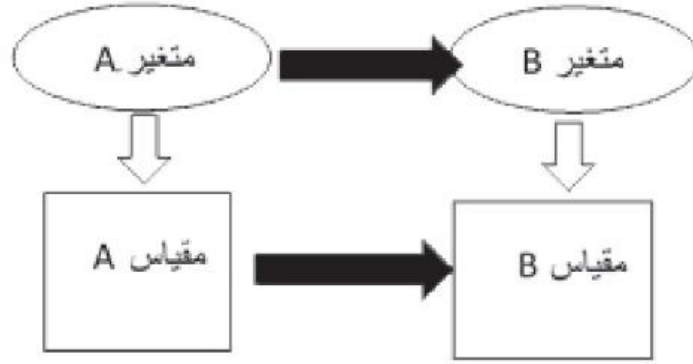
وغالبًا ما تسمى الظاهرة الأساسية أو المفهوم البنائي الأساسي المراد أن يعكسه المقياس بالمتغير الكامن (latent variable). ما هو المتغير الكامن بالضبط؟ يكشف مسمى المتغير الكامن عن اثنين من السمات الرئيسية. فبالنظر إلى المثال الخاص بتطلعات الآباء لما سينجزه الأطفال. أولاً، إن هذه التطلعات كامنة وليست ظاهرة أو واضحة. فطموحات الآباء لإنجازات أطفالهم لا يمكن ملاحظتها بصورة مباشرة. بالإضافة إلى ذلك، فإن المفهوم البنائي الخاص بطموحات وتطلعات الآباء لإنجازات أطفالهم هو متغير وليس ثابت - أي إن بعض جوانبه متغيرة، مثل قوته (شدته)، أو مقداره (أهميته). فقد تختلف تطلعات الآباء بشأن إنجازات أطفالهم وفقاً للوقت (على سبيل المثال، خلال مرحلة طفولتهم مقابل مرحلة المراهقة)، والمكان (على سبيل المثال، في مجال رياضي مقابل الفصل الدراسي)، أو أشخاص (مثل الآباء الذين تختلف خلفياتهم أو وظائفهم)، أو أي مزيج من هذه الأبعاد وغيرها. والمتغير الكامن هو الظاهرة الفعلية التي تكون موضع الاهتمام - في هذه الحالة، هو تطلعات الآباء لإنجازات الطفل.

وثمة جانب آخر جدير بالملاحظة في المتغير الكامن، هو أنه عادةً ما يكون سمة من سمات الفرد الذي هو مصدر البيانات. وهكذا، في مثالنا الحالي، فإن تطلعات الآباء هي سمة من سمات الآباء وليس الأطفال. ووفقاً لذلك، فإننا نقوم بتقييمه من خلال جمع بيانات حول معتقدات الآباء من الآباء أنفسهم. في حين أنه قد تكون هناك ظروف يكون فيها شكل ما من أشكال إعطاء المعلومات بالوكالة أو النيابة عن الغير (على سبيل المثال، مطالبة الوالدين بإعطاء معلومات عن بعض خصائص أطفالهم) أمراً مناسباً. بشكل عام، سنطلب من المستجيبين أن يقرروا ذاتياً المعلومات المتعلقة بأنفسهم. وعندما لا يكون الأمر كذلك، كما هو الحال في دراسة تشمل الآباء الذين يصفون تطلعات أطفالهم لأنفسهم، يجب توخي الحذر عند تفسير المعلومات الناتجة عن ذلك.

ويمكن القول، في هذه الحالة الافتراضية، بأن المتغير الكامن يمكن وصفه بشكل أكثر دقة على أنه تصورات الآباء لتطلعات أطفالهم وليس تطلعات الأطفال في حد ذاتها. وبالمثل، إذا طلبنا من مجموعة من المتسوقين تقييم خصائص متجر معين، فإننا نقوم بتقييم تصورات المتسوقين بدلاً من جوانب المتجر نفسه (والتي يمكن تقييمها بسهولة أكبر عن طريق الملاحظة المباشرة). إن درجة أهمية التمييز بين تقييم تصورات المستجيبين فيما يتعلق ببعض المحفزات الخارجية (على سبيل المثال، التصورات المتعلقة بالمتجر)، مقارنةً بخصائص المحفزات الخارجية (على سبيل



المثال، المتجر نفسه)، سوف تعتمد على الظروف والأهداف المحددة للتقييم. ومع ذلك، وفي جميع الحالات، من المهم مراعاة هذا التمييز وتقديم تفسيرات مناسبة للبيانات الناتجة عن ذلك.



شكل (2-1): تتطابق العلاقات بين الأدوات (المقاييس) مع العلاقات بين المتغيرات الكامنة فقط عندما يتطابق كل مقياس مع المتغير الكامن الخاص به

وعلى الرغم من أننا لا نستطيع أن نلاحظه أو نحدده كمياً بشكل مباشر، فمن المفترض أن المتغير الكامن يأخذ قيمة محددة في إطار مجموعة محددة من الشروط. والمقصود من المقياس الذي تم تطويره لقياس متغير كامن هو تقدير مقداره الفعلي في وقت ومكان القياس لكل شيء يتم قياسه. هذا المقدار الفعلي الذي لا يمكن ملاحظته هو الدرجة الحقيقية.

المتغير الكامن كسبب مفترض لقيم البند:

ينطوي مفهوم المتغير الكامن على وجود علاقة معينة بينه وبين البنود التي تعبر عنه. ويُعتبر المتغير الكامن سبباً في درجة البند - أي قوة أو مقدار المتغير الكامن (أي قيمة درجاته الحقيقية) يُفترض أنها تتسبب في البند (أو مجموعة من البنود) للحصول على قيمة معينة.

ومن الأمثلة التي قد تعزز هذه النقطة: فيما يلي مجموعة بنود افتراضية لتقييم تطلعات الآباء لإنجازات الطفل:

1- إنجازات طفلي تحدد نجاحي.

2- سأفعل أي شيء تقريباً لضمان نجاح طفلي.

3- ليست تضحية كبيرة إذا ساعدت طفلي على تحقيق النجاح.

#### 4- إنجازات طفلي أكثر أهمية بالنسبة لي من أي شيء آخر يمكنني التفكير فيه.

إذا تم إعطاء الفرصة للآباء للتعبير عن مدى موافقتهم القوية على كل بند من هذه البنود، فإن تطلعاتهم الكامنة لإنجازات الطفولة ينبغي أن تؤثر على استجاباتهم. وبعبارة أخرى، ينبغي لكل بند أن يبين مدى قوة المتغير الكامن (التطلعات لإنجازات الأطفال). والدرجات التي يتم الحصول عليها في البند هي نتيجة ناجمة عن قوة أو مقدار المتغير الكامن بالنسبة لذلك الشخص في ذلك الوقت بالذات.

وتتطوي العلاقة السببية بين متغير كامن ومقياس على علاقات تجريبية معينة. على سبيل المثال، إذا كانت قيمة البند ناتجة عن متغير كامن، فينبغي أن يكون هناك ارتباط بين هذه القيمة والدرجة الحقيقية للمتغير الكامن. وكنتيجة لارتباط كل مؤشر من المؤشرات بالمتغير الكامن، فإنه ينبغي أن ترتبط هذه المؤشرات أيضاً مع بعضها البعض. ونظراً لأننا لا نستطيع تقييم الدرجة الحقيقية بشكل مباشر، فإننا لا نستطيع حساب الارتباط بينها وبين البند. ومع ذلك، عندما ندرس مجموعة من البنود التي يُفترض أنها ناتجة عن نفس المتغير الكامن، فإنه يمكننا أن نفحص وندرس علاقاتها مع بعضها البعض. لذلك، إذا كانت لدينا عدة بنود مثل تلك المذكورة أعلاه والتي تقيس طموحات الآباء لإنجازات الأطفال، يمكننا أن ننظر مباشرة في كيفية ارتباطها مع بعضها البعض، ونستدعي المتغير الكامن كأساس للارتباطات بين البنود، وأن نستخدم هذه المعلومات للاستنتاج والاستدلال على مدى ارتباط كل بند ارتباطاً قوياً بالمتغير الكامن. وسأشرح بعد قليل كيف يمكن أن نتعلم ونستفيد من كل هذا من خلال الارتباطات المتبادلة بين البنود. أولاً، ومع ذلك، سوف أقوم بعرض بعض الإجراءات التخطيطية للمساعدة في جعل هذا التفسير أكثر وضوحاً.

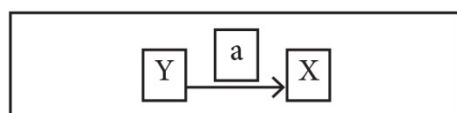
#### مخططات المسار:

ستقتصر تغطيه هذا الموضوع على استعراض موجز للقضايا ذات الصلة بتطوير المقياس. ولمزيد من التعمق في هذا الموضوع، يمكن الرجوع إلى «أشير» (Asher, 1983) أو «لوهلين» (Loehlin, 1998).

#### الاصطلاحات التخطيطية:

تعد مخططات المسار طريقة لوصف العلاقات السببية بين المتغيرات. وعلى الرغم من أنه يمكن استخدامها جنباً إلى جنب مع تحليل المسار، وهو أسلوب تحليلي للبيانات، إلا أن مخططات المسار لها فائدة عامة أكبر كوسيلة لتحديد كيفية ارتباط مجموعة من المتغيرات ببعضها البعض. وتتقيد هذه المخططات ببعض الاصطلاحات. حيث يشير السهم المستقيم المرسوم من اسم متغير إلى اسم متغير آخر إلى أن المتغيرين الاثنان مرتبطان سببياً وأن اتجاه السببية هو ما يشير إليه السهم. وهكذا، فإن  $(Y \rightarrow X)$  يشير صراحة إلى أن المتغير  $(X)$  هو سبب في حدوث المتغير  $(Y)$ . وفي كثير من الأحيان يتم تحديد المسارات الترابطية بواسطة علامات أو حروف، مثل الحرف (a) في الشكل رقم (2-2).

وغياب سهم ما له أيضاً معنى صريح، ألا وهو أن اثنين من المتغيرات لا توجد علاقة بينهما. وبالتالي، فإن  $(E \rightarrow C \ D \rightarrow B \rightarrow A)$  تعني أن  $(A)$  يسبب  $(B)$ ، و  $(B)$  يسبب  $(C)$ ، وأن  $(C)$  و  $(D)$  لا توجد علاقة بينهما، وأن  $(D)$  يسبب  $(E)$ .



شكل (2-2): المسار السببي من المتغير  $(X)$  إلى المتغير  $(Y)$

وهناك اصطلاح آخر لمخططات المسار وهو طريقة تمثيل الخطأ (error)، والتي يتم تصويرها عادة كمتغير سببي إضافي. ومصطلح الخطأ هذا عبارة عن باقي (residual)، ويمثل كل مصادر التباين غير المحسوبة لأسباب أخرى مبينة بوضوح في الرسم التخطيطي.

ونظراً لأن مصطلح الخطأ هذا متبقي، فإنه يمثل الاختلاف بين القيمة الفعلية للمتغير  $(Y)$  والقيمة التي نتوقعها للمتغير  $(Y)$  والمستندة إلى المعرفة بالمتغيرات  $(X)$  و  $(Z)$  (في هذه الحالة؛ انظر الشكل 2-3). وفي بعض الأحيان، يتم افتراض مصطلح الخطأ، وبالتالي، لا يتم تضمينه في الرسم التخطيطي.

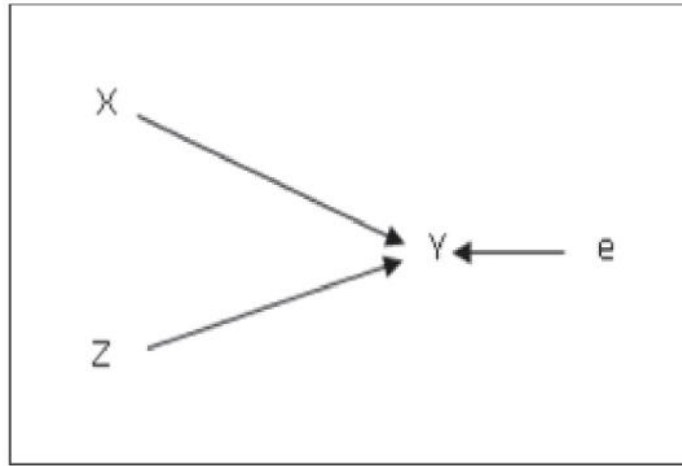
مخططات المسار في تطوير المقياس:

يمكن أن تساعدنا مخططات المسار في معرفه كيفية أو مدى ارتباط بنود المقياس بالمتغير الكامن سببياً. كما أنها يمكن أن تساعدنا أيضاً على فهم كيف أن بعض العلاقات بين البنود تنطوي على علاقات معينه بين البنود والمتغير الكامن. نبدأ بدراسة قاعدة حسابه بسيطة لمخططات المسار. دعنا نلقي نظرة على الرسم التخطيطي للمسار البسيط في الشكل رقم (2-4).

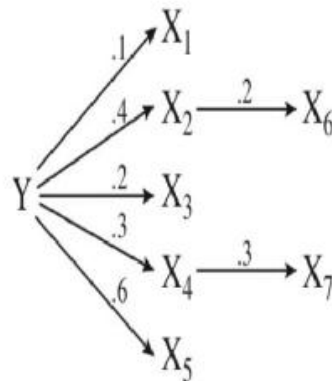
الأرقام على طول المسارات هي معاملات مسار معيارية (قياسية). وكل معامل مسار معياري يعبر عن قوه العلاقة السببية بين المتغيرات المرتبطة بالسهم. إن حقيقة كون المعاملات معيارية يعني أنها جميعاً تستخدم نفس المقياس (وحدات القياس) لقياس مقدار العلاقات السببية. في هذا الرسم التخطيطي، فإن المتغير (Y) هو سبب لكل من المتغيرات من (X1) إلى (X5). وتوجد علاقة هامة بين قيم معاملات المسار والارتباطات بين البنود (Xs) (التي تمثل البنود في حاله الرسم التخطيطي لمسار تطوير المقياس). بالنسبة للمخططات مثل هذه والتي يكون لها أصل مشترك واحد فقط (Y في هذه الحالة)، فإن الارتباط بين أي اثنين من البنود (Xs) تساوي ناتج ضرب معاملات الأسهم التي تشكل مساراً، من خلال (Y)، بين المتغيرات (X) محل الاهتمام. على سبيل المثال، يتم حساب الارتباط بين (X1) و (X5) عن طريق ضرب معاملي المسارين المعياريين اللذين يربطهما ب (Y). وهكذا، فإن معامل الارتباط بينهما يساوي  $(r_{1,5} = 0.6 \times 0.1 = 0.06)$ . كما تشترك المتغيرات (X6) و (X7) في (Y) كمصدر مشترك، لكن المسار الذي يربطهما يكون أطول. ومع ذلك، فإن قاعدة الحساب لا تزال سارية وتنطبق في هذه الحالة. فإذا بدأنا من (X7)، فإنه يمكننا تتبع العودة إلى (Y) ومن ثم التوجه إلى الأمام مرة أخرى إلى (X6) (أو في الاتجاه الآخر، من X6 إلى X7). وتكون النتيجة على النحو التالي:  $(0.2 \times 0.4 \times 0.3 = 0.0072)$ . وهكذا تكون قيمة  $(r_{6,7})$  تساوي 0,0072.

هذه العلاقة بين معاملات المسار والارتباطات توفر أساساً لتقدير المسارات بين متغير كامن والبنود التي يؤثر عليها. وعلى الرغم من أن المتغير الكامن هو متغير افتراضي وغير قابل للقياس، إلا أن البنود حقيقية ويمكن حساب الارتباطات بينها بشكل مباشر. وباستخدام هذه الارتباطات، والقاعدة البسيطة للحساب التي تمت مناقشتها للتو، وبعض الافتراضات حول العلاقات بين البنود والدرجات الحقيقية، يمكننا التوصل إلى تقديرات للمسارات بين البنود والمتغير الكامن. ويمكننا أن نبدأ بمجموعة من الارتباطات بين المتغيرات. وبعد ذلك، بالعمل إلى الخلف أي العكس

بأن نبدأ من العلاقة بين المسارات والارتباطات، يمكننا تحديد ما يجب أن تكون عليه قيم مسارات معينة إذا كانت الافتراضات صحيحة. دعونا ننظر في المثال الوارد في الشكل رقم (2-5).



شكل (2-3): متغيران بالإضافة إلى تحديد الخطأ للمتغير (Y)

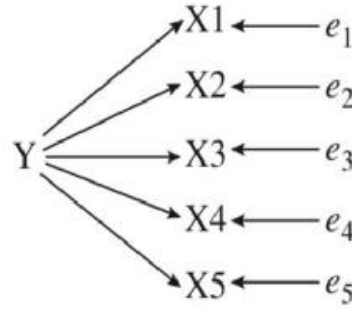


شكل (2-4): مخطط مسار ذو معاملات مسار والذي يمكن استخدامه لحساب الارتباطات بين المتغيرات

يشبه هذا الرسم التخطيطي المثال الذي تم تناوله سابقاً في الشكل رقم (2-4) باستثناء أنه لا توجد قيم مسارات، وتم حذف وإسقاط المتغيرين (X6) و (X7)، وتمثل المتغيرات (X) المتبقية بنود المقياس، ولكل بند متغير (خطأ error) يؤثر فيه بخلاف المتغير (Y). ومتغيرات الخطأ (e) هذه هي متغيرات فريدة من نوعها في حالة كل بند وتمثل التباين المتبقي في كل بند لم يتم تفسيره بواسطة المتغير (Y). ويشير هذا المخطط إلى أن كافة البنود تتأثر بالمتغير (Y). بالإضافة

إلى ذلك، يتأثر كل منها بمجموعه فريدة من المتغيرات بخلاف المتغير (Y) والتي يتم التعامل معها بشكل جماعي على أنها خطأ (error).

ويمثل هذا المخطط المعد لكيفية ارتباط خمسة بنود كل على حدة بمتغير كامن واحد (Y). وتشير الأرقام العددية المعطاة لمتغيرات الخطأ (es) وبنود المقياس (Xs) إلى أن البنود الخمسة مختلفة وأن المصادر الخمسة للخطأ، واحد لكل بند، هي أيضاً مختلفة. ولا يحتوي المخطط على أسهم للانتقال مباشرة من بند (X) إلى بند آخر (X) أو الانتقال من متغير خطأ (e) إلى متغير خطأ آخر (e) أو من متغير خطأ (e) إلى بند (X) غير تلك المرتبطة بها. وتمثل هذه الجوانب من المخطط الافتراضات التي ستتم مناقشتها لاحقاً.



شكل (2-5): مخطط مسار يتضمن متغيرات الخطأ

إذا كانت لدينا خمسة بنود فعلية أكملت الإجابة عليها مجموعة من الأشخاص، فستكون لدينا درجات بند يمكن أن نربطها مع بعضها بعد ذلك. ولقد سمحت القاعدة التي تمت دراستها في وقت سابق بحساب الارتباطات من معاملات المسار. ومع إضافة بعض الافتراضات، فإنه يتيح لنا أيضاً حساب معاملات المسار من الارتباطات، أي إنه يمكن استخدام الارتباطات المحسوبة من البنود الفعلية لتحديد مدى ارتباط كل بند بالمتغير الكامن. على سبيل المثال، إذا كان الارتباط بين (X1) و(X4) يساوي (0,49)، فعندئذ نعرف أن ناتج قيم المسار الذي يمتد من (Y) إلى (X1) والمسار الذي يمتد من (Y) إلى (X4) يساوي (0,49). ونحن نعرف هذا لأن قاعدتنا الحسابية أثبتت أن الارتباط بين متغيرين يساوي ناتج ضرب معاملات المسار على طول المسار الذي يربطهما معاً. وإذا افترضنا أيضاً أن قيمتي المسار متساوية، فيجب أن يكون كلاهما (0,70)(1).

مزيد من التوسع في نموذج القياس:

## افتراضات القياس الكلاسيكية (التقليدية):

إن نموذج القياس الكلاسيكي أو التقليدي - الذي يؤكد أن الدرجة الملاحظة ( $X$ )، الناتجة عن حاصل جمع الدرجات الحقيقية<sup>2</sup> ( $T$ )، بالإضافة إلى خطأ القياس ( $e$ ) - تبدأ بافتراضات عامة حول البنود وعلاقاتها بالمتغير الكامن ومصادر الخطأ، وهذه الافتراضات هي:

١- يختلف مقدار الخطأ المرتبط بكل بند من البنود على حدة بشكل عشوائي. ويكون للخطأ المرتبط بكل بند من البنود على حدة متوسط يساوي (صفرًا) عند تجميعه عبر عدد كبير من الأشخاص. وبالتالي، فإن متوسطات البنود تميل إلى عدم التأثر بالخطأ عندما يكمل عدد كبير من المجيبين الإجابة على البنود.

٢- لا يرتبط متغير الخطأ (error term) لأي بند مع متغير الخطأ لبند آخر. أما المسارات الوحيدة التي تربط البنود فهي تمر دائماً عبر المتغير الكامن، ولا تمر أبداً عبر أي متغير للخطأ.

٣- لا ترتبط متغيرات الخطأ مع الدرجة الحقيقية للمتغير الكامن. لاحظ أن المسارات المنبثقة من المتغير الكامن لا تمتد للخارج إلى متغير الخطأ. ويستهدف السهم بين بند ما ومتغير الخطأ الخاص الاتجاه الآخر.

الافتراضات الأولى والثاني أعلاه هما افتراضين إحصائيين شائعين تكمنان أو تقفان وراء العديد من الإجراءات التحليلية. أما الافتراض الثالث فهو تعريف “الخطأ” على أنه الباقي أو المتبقي بعد الأخذ في الاعتبار كل العلاقات بين مجموعه من المنبآت والنتيجة أو، في هذه الحالة، مجموعه من البنود ومتغيرها الكامن.

## الاختبارات المتكافئة (المتوازية):

تعتمد نظرية القياس الكلاسيكية، في شكلها الأكثر تقليدية، على افتراض اختبارات التكافؤ أو التوازي parallel tests. وينبع مصطلح “اختبارات التكافؤ” من حقيقة أنه يمكن للمرء أن ينظر إلى كل بند على حدة على أنه “اختبار” لقيمة المتغير الكامن. ولأغراضنا، فإن الإشارة إلى البنود المتكافئة ستكون أكثر دقة. ومع ذلك، سوف أؤجل استخدام ذلك وأستخدم الاسم التقليدي.



ومن مزايا نموذج اختبارات التكافؤ أن افتراضاته تجعل من السهل للغاية التوصل إلى استنتاجات مفيدة حول كيفية ارتباط كل بندين من البنود على حدة بالمتغير الكامن بناءً على ملاحظتنا عن كيفية ارتباط البنود ببعضها البعض. في وقت سابق، اقترحت أنه بمعرفة الارتباطات بين البنود وفي ظل افتراضات معينة، يمكن للمرء عمل استدلالات واستنتاجات بشأن المسارات المؤدية من متغير سببي إلى بند ما. كما هو مبين في الفصل التالي، فإن القدرة على تعيين وتحديد قيمة رقمية للعلاقات بين المتغير الكامن والبنود نفسها مهمة جداً. ومن ثم فإنني، في هذا القسم، سأبحث بشيء من التفصيل كيف تؤدي افتراضات اختبارات التكافؤ إلى استنتاجات معينة تجعل ذلك ممكناً.

والأساس المنطقي الذي يقوم عليه نموذج الاختبارات المتكافئة هو أن كل بند من بنود المقياس هو على وجه التحديد مقياس جيد للمتغير الكامن مثل أي بند آخر من بنود المقياس. وبالتالي، فإن كل بند من البنود تكون متكافئة تماماً، أي إن علاقة كل بند بالمتغير الكامن يُفترض أنها متطابقة مع علاقة كل بند آخر بهذا المتغير ويُفترض أيضاً أن يكون مقدار الخطأ الموجود في كل بند متطابقاً أيضاً. وبشكل تخطيطي، يمكن تمثيل هذا النموذج على النحو المبين في الشكل رقم (٦-٢).

ويضيف هذا النموذج افتراضين إلى تلك الافتراضات المذكورة آنفاً وهما:

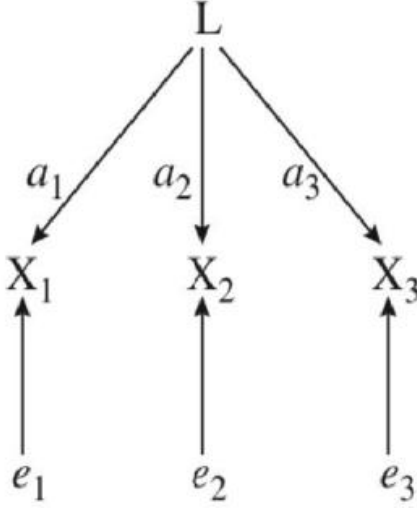
١- يفترض أن يكون مقدار التأثير من المتغير الكامن لكل بند هو نفسه بالنسبة لجميع البنود.

٢- يفترض أن يحتوي كل بند نفس مقدار الخطأ لأي بند آخر، مما يعني أن تأثير العوامل الأخرى غير المتغير الكامن متساوٍ لكافة البنود.

تعني هذه الافتراضات المضافة أن ارتباطات كل بند والدرجة الحقيقية تكون متطابقة. والقدرة على التأكيد بأن هذه العلاقات متساوية هو أمر مهم لأنها تؤدي إلى وجود وسيلة لتحديد قيمه كل من هذه الارتباطات المتطابقة أو المتماثلة، وهذا بدوره يؤدي إلى طريقة لقياس مدى الثبات، والذي ستنم مناقشتها في الفصل التالي.

إن التأكيد على أن الارتباطات بين الدرجة الحقيقية وكل بند متساوية يتطلب كل الافتراضات السابقة. ومربع الارتباط هو نسبة التباين المشترك بين متغيرين. لذلك، إذا كانت الارتباطات بين

الدرجة الحقيقية وكل بند من بندين متساوية، فإن نسب التباين المشترك بين الدرجة الحقيقية وكل بند يجب أن تكون متساوية أيضاً. افترض أن الدرجة الحقيقية تسهم بنفس مقدار التباين لكل بند من البندين. فإن هذا المقدار يمكن أن يكون نسبه متساوية من التباين الكلي لكل بند فقط إذا كانت البنود لها تباينات كلية متطابقة. ولكي تكون التباينات الكلية متساوية للبندين، فإنه يجب أن يكون مقدار التباين الذي يتلقاه كل بند من مصادر أخرى غير الدرجة الحقيقية متساوياً أيضاً. ونظراً لأن كل مصادر التباين بخلاف الدرجة الحقيقية يتم تجميعها معاً أنها خطأ، فإن هذا يعني أنه يجب أن يكون للبندين تباينات خطأ متساوية. على سبيل المثال، إذا حصلت (X1) على تسع وحدات عشوائية من التباين من درجتها الحقيقية ووحدة واحدة من الخطأ، فإن نسبة الدرجة الحقيقية ستكون (٩٠٪) من التباين الكلي. إذا حصلت (X2) أيضاً على تسع وحدات تباين من الدرجة الحقيقية، فإن هذه الوحدات التسع يمكن أن تكون (٩٠٪) من الإجمالي فقط إذا كان التباين الكلي يساوي (١٠). ولا يمكن أن يساوي المجموع القيمة (١٠) إلا إذا ساهم الخطأ بوحدة واحدة في (X2) مثلما ساهمت في (X1). عندئذٍ، فإن الارتباط بين كل بند والدرجة الحقيقية يساوي الجذر التربيعي لنسبة تباين كل بند الذي يعزى إلى الدرجة الحقيقية، أو ما يقرب من (٠,٩٥) في هذه الحالة.



شكل (٢-٦): مخطط لنموذج الاختبارات المتكافئة، حيث تكون جميع المسارات من المتغير الكامن (L) إلى البنود (X1, X2, X3) متساوية في القيمة مع بعضها البعض، وكذلك جميع المسارات من متغيرات الخطأ إلى البنود

وبالتالي، نظرًا لأن نموذج الاختبارات المتكافئة يفترض أن مقدار التأثير من المتغير الكامن هو نفسه لكل بند وأن المقدار من المصادر الأخرى (الخطأ) هي نفسها لكل بند، فإن نسب تباين البند المنسوبة إلى متغير كامن والخطأ يكونان متساويان بالنسبة لجميع البنود. وهذا يعني أيضًا أنه وفقًا لافتراضات الاختبارات المتكافئة، تكون معاملات المسار المعيارية من المتغير الكامن لكل بند متساوية لكافة البنود. وكان من المفترض أن تكون معاملات المسار المعيارية متساوية مما أتاح لنا إمكانية، في مثال سابق، حساب معاملات المسار من الارتباطات بين البنود. يجب أن تساعدنا قاعدة مخطط المسار المتعلقة بمعاملات المسار مع الارتباطات، والتي تمت مناقشتها سابقًا، في فهم سبب استمرار وجود هذه المعادلة أو التكافؤ عندما يقبل المرء الافتراضات السابقة.

كما أن افتراضات هذا النموذج تعني ضمناً أن الارتباطات بين البنود تكون متطابقة أو متماثلة (على سبيل المثال، الارتباط بين (X1) و (X2) مطابق للارتباط بين (X1) و (X3) أو بين (X2) و (X3). كيف نتوصل إلى هذا الاستنتاج من الافتراضات؟ الارتباطات كلها متماثلة لأن الآلية الوحيدة لحساب الارتباط بين أي بندين هي المسار من خلال المتغير الكامن الذي يربط بين

هذين البندين. على سبيل المثال، يتم ربط  $(X1)$  و  $(X2)$  فقط بواسطة المسار المكون من المسارين  $(a1)$  و  $(a2)$ . ويمكن حساب الارتباط عن طريق تتبع المسار الذي يربط بين البندين المعنيين وحاصل ضرب قيم المسار. بالنسبة لأي بندين من البنود، يستلزم ذلك ضرب مسارين لهما قيم متطابقة (أي،  $a1 = a2 = a3$ ). وبطبيعة الحال، ستكون الارتباطات المحسوبة من خلال ضرب القيم المتساوية متساوية.

كما ان الافتراضات تعني أن كل من هذه الارتباطات بين البنود تساوي مربعاً أي مساراً من المتغير الكامن إلى بند من البنود على حدة. كيف نصل إلى هذا الاستنتاج؟ حاصل ناتج ضرب مسارين مختلفين (على سبيل المثال،  $a1$  و  $a2$ ) يكون مطابقاً لمربع أي من المسارين لأن معاملات كلا المسارين متماثلان. إذا كانت  $a1 = a2 = a3$ ، وكان حاصل ضرب  $(a2 \times a1) =$  حاصل ضرب  $(a1 \times a3) =$  حاصل ضرب  $(a2 \times a3)$ ، فعندئذٍ يجب أن يساوي كل من هذه النواتج الأخيرة أيضاً قيمة أي من المسارات  $(a)$  مضروباً في نفسه أي مربع المسار  $(a)$ . وقد يؤدي النظر إلى الوراء في الشكل رقم (٦-٢) إلى توضيح هذه العلاقات والآثار المترتبة عليها بشكل أكبر.

ويستنتج أيضاً من افتراضات هذا النموذج أن نسبة الخطأ المرتبطة بكل بند هو مكمل لنسبة التباين المرتبطة بالمتغير الكامن. وبعبارة أخرى، فإن أي تأثير على بند معين لا يتم تفسيره بواسطة المتغير الكامن يجب أن يتم تفسيره بواسطة متغير الخطأ (error). ويفسر ويشرح هذين التأثيرين معاً (١٠٠٪) من التباين في أي بند معين. وهذا ببساطة لأنه يتم تعريف متغير الخطأ  $(e)$  على أنه يشمل جميع مصادر التباين في البند بخلاف المتغير الكامن.

وتدعم هذه الافتراضات استنتاجاً واحداً آخر على الأقل؛ نظراً لأن كل بند يتأثر بنفس المقدار بالمتغير الكامن كما أن تأثير كل متغير من متغيرات الخطأ على البند المقابل له متساوٍ أيضاً، فإن كافة البنود تكون لها متوسطات متساوية وتباينات متساوية. إذا كان المصدران الوحيدان اللذان يمكن أن يؤثرًا على المتوسطات هما مصدران متماثلان أو متطابقان بالنسبة لكافة البنود، فمن الواضح أن المتوسطات الخاصة بكافة البنود ستكون متطابقة أيضاً. وينطبق هذا المنطق أيضاً على التباينات الخاصة بالبنود.

وفي الختام، فإن نموذج الاختبارات المتكافئة أو المتوازية يفترض ما يلي:

١- عشوائية الخطأ.

٢- عدم ارتباط الأخطاء مع بعضها البعض.

٣- الأخطاء غير مرتبطة بالدرجة الحقيقية.

٤- يؤثر المتغير الكامن على جميع البنود بنفس القدر.

٥- مقدار الخطأ لكل بند يكون متساوياً.

تسمح لنا هذه الافتراضات أن نتوصل إلى مجموعه متنوعة من الاستنتاجات المثيرة للاهتمام. وعلاوة على ذلك، يتيح لنا النموذج استخلاص استنتاجات حول المتغير الكامن بناءً على ارتباطات البنود مع بعضها البعض. ومع ذلك، فإن النموذج يحقق هذا العمل من خلال وضع افتراضات دقيقة وصارمة إلى حدٍ ما.

نماذج بديلة:

وكما يحدث، فإن جميع الافتراضات المقيدة بقوة والمرتبطة بالاختبارات المتكافئة أو المتوازية التامة ليست ضرورية من أجل تحقيق استنتاجات مفيدة حول علاقة الدرجات الحقيقية بالدرجات الملاحظة. إن النموذج القائم على ما يسمى تقنياً باختبارات «تاو» المتكافئة ( $\tau$ -equivalent) يجعل أحد الافتراضات أكثر تحدياً - وهو أن مقدار التباين في الخطأ المرتبط ببند معين لا يلزم أن يساوي تباين الخطأ في البنود الأخرى (على سبيل المثال، «الين، وين» Allen & Yen, 1979). ولكن لا تزال اختبارات «تاو» المتكافئة تتطلب درجات حقيقية متطابقة للبنود، على الرغم من أن التخفيف البسيط من هذا الافتراض يُعرف باسم اختبارات «تاو» المترادفة أو المتكافئة في الأساس essentially  $\tau$ -equivalent tests (أو في بعض الأحيان اختبارات متكافئة عشوائياً randomly parallel tests). قد يكون لأي زوج من البنود الخاضعة أو المقيدة باختبارات «تاو» المترادفة أو المتكافئة في الأساس درجات حقيقية تختلف باختلاف الثابت. وبطبيعة الحال، فإن إضافة ثابت إلى بند واحد لا يكون له تأثير على أي ارتباط ينطوي على ذلك البند، لأن الارتباطات عبارة عن مقادير معيارية. وبالتالي، فإن العلاقة بين أي زوج من البنود أو بين الدرجة الحقيقية للبند والدرجة التي حصل عليها البند لا تتأثر بالتخفيف من الافتراضات

الخاصة باختبارات «تاو» المتكافئة الصارمة مع تلك الافتراضات الخاصة بنموذج «تاو» المترادف أو المتكافئ في الأساس. إذن، ما تم قوله حتى الآن حول اختبارات «تاو» المتكافئة ينطبق أيضاً على نموذج «تاو» المترادف أو المتكافئ في الأساس. في أي من هاتين الحالتين، قد لا تكون القيم المعيارية للمسارات من المتغير الكامن لكل بند متساوية. ومع ذلك، فإن القيم غير المعيارية للمسار من المتغير الكامن إلى كل بند (أي، المقدار المقابل لنسبة التأثير التي يمتلكها المتغير الكامن على كل بند) يفترض أنها لا تزال متطابقة بالنسبة لكافة البنود. وهذا يعني أن البنود متكافئة أو متوازية فيما يتعلق بمقدار تأثيرها بالمتغير الكامن ولكنها لا تتأثر بالضرورة بنفس القدر من العوامل الخارجية التي يتم تجميعها معاً كخطأ. وفي ظل افتراضات التكافؤ التام، لا تقوم بنود مختلفة فقط بالتأثير على الدرجة الحقيقية بنفس القدر أو الدرجة؛ بل أيضاً يكون لمكونات الخطأ الخاصة بهم نفس التأثير. وتُعد مكافئ «تاو» (تاو) "tau" هي المقابل اليوناني لـ  $\tau$ ، كما هو الحال في الدرجة الحقيقية) أسهل بكثير في التعامل لأنه لا يفترض شرط "تساوي الأخطاء". ونظراً لأن الأخطاء قد تختلف، فإن متوسطات وتباينات البنود قد تختلف أيضاً. والافتراضات الأكثر تحريراً لهذا النموذج تكون جذابة لأن العثور على مقاييس متكافئة للتباين المتساوي هو أمر نادر الحدوث. ويتيح لنا هذا النموذج أن نتوصل إلى العديد من الاستنتاجات ذاتها التي توصلنا إليها كما في الاختبارات المتكافئة أو المتوازية التامة ولكن مع افتراضات أقل تقييداً. وقد يرغب القراء في مقارنة هذا النموذج بمناقشة «نونالي وبيرنستين» (Nunnally and Bernstein, 1994) لنموذج معاينة النطاق أو المجال.

يعتبر بعض مطوري المقاييس أن نموذج "تاو" المترادف أو المتكافئ في الأساس هو نموذج مقيد ومتشدد جداً. وفي نهاية المطاف، كم مرة يمكننا أن نفترض أن كل بند يتأثر بالمتغير الكامن بنفس الدرجة؟ الاختبارات التي تم تطويرها في إطار ما يسمى بالنموذج المتقارب أو التقاربي congeneric model (جوريسكوج، 1971 Jöreskog) تخضع لمجموعه من الافتراضات الأكثر مرونة "غير المشددة" (انظر كارمينز، وماكلفير 1981 McLver, & Carmines، لمناقشة الاختبارات المتقاربة أو التألفية). يفترض هذا النموذج (فيما يتجاوز افتراضات القياس الأساسية) بكل بساطة أن كافة البنود تشترك في متغير كامن عام. وأنها لا تحتاج إلى أن يكون لها علاقات قوية بنفس القدر مع المتغير الكامن، ولا ينبغي أن تكون تباينات الخطأ الخاصة بها متساوية. ويجب أن يفترض المرء فقط أن كل بند يعكس

الدرجة الحقيقية إلى حدٍ ما. وبطبيعة الحال، فإنه كلما كان بند يرتبط بقوة بالدرجة الحقيقية، كان المقياس أكثر ثباتاً.

المدخل الأقل تقييداً يتمثل في نموذج العامل العام general factor، الذي يتيح للمتغيرات الكامنة المتعددة أن تشكل الأساس لمجموعة معينة من البنود. وقد ناقش كل من “كارمينز وماكلفير” (Carmines and McLver, 1981)، و”لوهلين” (Loehlin, 1998)، و”لونج” (Long, 1983) مزايا هذا النوع من النماذج العامة للغاية، وأهمها تحسين التطابق مع البيانات في العالم الحقيقي. وغالباً ما تتضمن أساليب نمذجة المعادلة البنائية التحليل العاملي factor analyses في نماذج القياس الخاصة بها؛ وتُعد الحالات التي تشكل فيها المتغيرات الكامنة المتعددة مجموعة من المؤشرات مثلاً على نموذج العامل العام (لوهلين، 1998).

ويُعد النموذج المتقارب أو التقاربي هو حالة خاصة لنموذج العامل (أي حالة العامل المفرد أو الوحيد). وبالمثل، فإن مقياس نموذج “تاو” المترادف أو المتكافئ في الأساس هو حالة خاصة من المقياس المتقارب - الذي يفترض أن تكون العلاقات بين البنود والمتغير الكامن الخاص بها متساوية. وأخيراً، يمثل الاختبار المتكافئ التام حالة خاصة لنموذج «تاو» المترادف أو المتكافئ في الأساس، مما يضيف افتراض وجود علاقات متساوية بين كل بند ومصادر الخطأ المرتبطة به.

وهنا تجدر الإشارة إلى إستراتيجية قياس أخرى. هذه الإستراتيجية هي نظرية الاستجابة للبند أو المفردة (item response theory «IRT»). تم استخدام هذا المدخل في المقام الأول، ولكن ليس بشكل حصري، مع بنود الاستجابة الثنائية (على سبيل المثال، الصحيح في مقابل غير الصحيح) في تطوير اختبارات القدرة. وتفترض نظرية الاستجابة للبند أن لكل بند من البنود على حدة حساسية خاصة به للمتغير الكامن، ممثله بالمنحنيات المميزة للبند - وهو رسم بياني للعلاقة بين قيمة المتغير الكامن (على سبيل المثال، القدرة) واحتمال استجابة معينة لأحد البنود (على سبيل المثال، الإجابة عليه بشكل صحيح). وبالتالي، يكشف المنحنى المميز للبند عن مقدار القدرة التي يتطلبها بند ما لتتم الإجابة عليه بشكل صحيح. سنتناول بشكل أكبر نظرية الاستجابة للبند في الفصل السابع.

فيما عدا دراسة نظرية الاستجابة للبند (IRT) في الفصل السابع ومناقشة التحليل العاملي في الفصل السادس، سوف نركز بشكل أساسي على النماذج المتكافئة أو المتوازية ونموذج “تاو”

المترادف أو المتكافئ في الأساس وذلك لعدة أسباب. أولاً، أنها تجسد وتمثل نظرية القياس «الكلاسيكية». ثانياً، أن مناقشة الآليات التي تعمل بها النماذج الأخرى بصورة سريعة يمكن أن تكون أمراً مرهقاً. وأخيراً، أثبتت النماذج الكلاسيكية أنها مفيدة جداً للعلماء الاجتماعيين الذين لديهم اهتمامات أساسية غير القياس، ومع ذلك، فإنهم يأخذون القياس الدقيق على محمل الجد. وهذه المجموعة من الباحثين هي الجمهور الذي تمت كتابة هذا الكتاب له. وبالنسبة لهؤلاء الأفراد، فإن إجراءات تطوير المقياس التي تتبع النموذج الكلاسيكي تسفر عموماً عن مقاييس مرضية. وفي الواقع، وعلى حد علمي أنه على الرغم من عدم وجود حيلة متاحة بسهولة، أعتقد أنه (اختبار القدرة الخارجية) قد تم تطوير الغالبية العظمى من المقاييس المعروفة والمُعترف بها بدرجة كبيرة والمستخدمة في أبحاث العلوم الاجتماعية باستخدام مثل هذه الإجراءات.

### اختيار نموذج سبي:

إن اختيار النموذج السبي الذي يشكل أساساً لمتغير ما، عندما يكون ذلك ممكناً، يمكن أن يكون جانباً مهماً من جوانب القياس. يمكن أحياناً تكييف مفهوم المتغير ذاته بدقة في بداية مشروع بحثي لجعل قياسه في النهاية أكثر سهولة وقابلية للتطبيق. على سبيل المثال، لنأخذ في الاعتبار أن باحثاً يريد تقييم كيفية تأثير بيئة العمل المادية على إنتاجية الموظف. قد يكون أحد الأساليب هو وضع قائمة طويلة من العوامل البيئية التي يُعتقد أنها تؤثر على الإنتاجية - مثل الإضاءة أو الإحساس بالخصوصية أو الوصول إلى جهاز حاسب آلي - وتطوير أداة تجعل العاملين يحددون مدى وجود هذه العوامل في مكان عمل معين. تتمثل مشكلة هذا المدخل أو الأسلوب في أن الأداة قد تكون في النهاية عبارة عن دليل index وليست مقياساً أو ربما مزيجاً من الاثنين. أي إن المؤشرات (مثلاً، الإضاءة الجيدة والخصوصية المعقولة والوصول إلى جهاز حاسب آلي) قد لا تشترك في الواقع في مسألة مشتركة بل قد يكون لها تأثير مشترك، ألا وهو تحسين بيئة العمل. وبدلاً من ذلك، إذا نظر الباحث في مشكلة القياس المحتملة في وقت مبكر من عملية البحث، فإنه ربما يكون قد قرر تحديد أو وضع مفهوم للمتغير بطريقة مختلفة إلى حد ما. فعلى سبيل المثال، إذا كان الباحث قد عرّف المتغير محل الاهتمام على أنه تصورات الموظفين لبيئة العمل، فإن هذا التعريف قد يؤدي إلى مجموعة أكثر مرونة من البنود. على سبيل المثال، يمكن أن يُطلب من الموظفين الموافقة على بنود مثل: “بيئة مكان العمل الخاص بتوفر الأجهزة الأساسية التي أحتاجها للقيام بعملتي بشكل فعال”. هنا، لا يكون المتغير الكامن سمه من سمات البيئة في حد ذاتها ولكنه



تصورات الموظفين. إن الطريقة التي يدرك بها الموظفون البيئة على أنها السبب العام الذي يقود ردودهم على كل بند من البنود على حدة، أجد أنه من الأسهل افتراض أن الموظف لديه شعور ببيئة العمل التي ستؤدي إلى إجابات عبر مجموعه من الأسئلة حول مدى كفايتها بدلاً من تصور البيئة نفسها كسبب لإجابات الموظف. وعلاوة على ذلك، فإن الطبيعة النفسية لتصورات الموظفين قد تكون في الواقع أقرب إلى ما يعتبره الباحث ذا صلة بالإنتاجية من مجرد وجود أو عدم وجود سمات بيئية محددة. أي، إذا كان العامل ينظر إلى البيئة على أنها مواتية للإنتاجية قد يكون متغيراً أكثر أهميه من حكم شخص آخر فيما يتعلق بمدى كفاية بيئة العمل. ولذلك؛ فإن وضع تصور للمتغير محل الاهتمام بهذه الطريقة قد يخدم السؤال البحثي الأساسي بشكل جيد، مع احتمال تسهيل القياس النهائي للمتغير أيضاً.

وبطبيعة الحال، إذا لم يكن المتغير ذاته يقدم تصوراً سببياً يتفق مع إستراتيجية القياس المباشرة، فلا ينبغي المساس بسلامة المتغير موضع الاهتمام. قد تساعد بعض المداخل والأساليب الباحث في التغلب على القيود المتأصلة في المتغير والطريقة التي يتم بها قياسه إجرائياً. ولكن إذا أمكن تبسيط تصور بديل مقبول للمتغير والنموذج الذي يربطه بمؤشرات، فقد يكون من الممكن تطوير أداة قياس تلبي وتفي بمجموعة أبسط من الافتراضات، ومن ثم يمكن استكشافها باستخدام أدوات تحليلية أقل تعقيداً. ومن المؤكد أن امتلاك الأدوات اللازمة للتعامل مع المواقف والحالات الأكثر تعقيداً يُعد أمراً جيداً، ولكن تجنب تلك التعقيدات واستبعاد الحاجة إلى هذه الأدوات الأكثر تقدماً قد يكون أفضل، على افتراض أنه يحقق العدالة في المفهوم البنائي.

تمارين:

١- كيف يمكننا استنتاج العلاقة بين المتغير الكامن والبندين المرتبطين به على أساس الارتباطات بين البندين؟

٢- ما هو الفرق الرئيسي في الافتراضات بين الاختبارات المتكافئة ونموذج «تاو» المترادف أو المتكافئ في الأساس؟

٣- ما هو نموذج القياس الذي يفترض، إلى جانب الافتراضات الأساسية الشائعة لجميع طرق القياس، أن البنود تشترك في متغير كامن مشترك؟

٤- افترض أن نموذج «تاو» المترادف أو المتكافئ في الأساس مع درجة حقيقة T ومؤشرات A و B و C. في مثل هذا النموذج، يجب أن يكون لأي مؤشرين (على سبيل المثال، A و B) والذين يشتركان في درجة حقيقة لهما تغاير مطابق للتغاير بين زوجين آخرين من المؤشرات (على سبيل المثال، B و C) يشتركان في تلك الدرجة الحقيقية. ومع ذلك، لا يجب أن تكون الارتباطات بين أزواج المؤشرات المختلفة متساوية. اشرح سبب حدوث ذلك.

#### ملاحظات:

١- على الرغم من أن (-٠,٧٠) هو أيضاً الجذر التربيعي المسموح به للقيمة (٠,٤٩)، إلا أن اتخاذ القرار بين الجذر الموجب أو السالب عادةً ما يكون أقل أهمية مما يعتده المرء. طالما أن جميع البنود يمكن أن تكون مرتبطة بشكل إيجابي مع بعضها البعض (إذا لزم الأمر، من خلال عكس درجات بنود معينة، كما هو موضح في الفصل الخامس)، فإن إشارات معاملات المسار من المتغير الكامن لكل بند من البنود على حدة سوف تكون هي نفسها. لاحظ، مع ذلك، أن إعطاء إشارات إيجابية لهذه المسارات يعني ضمناً أن البنود تشير إلى المفهوم البنائي بشكل أكبر، في حين أن المعاملات السلبية تنطوي على عكس ذلك.

## الفصل الثالث

### الثبات (الموثوقية)

يُعد الثبات أو الموثوقية قضية أساسية في القياس النفسي. وأهمية الثبات تتضح بمجرد فهم معناه بشكل كامل وجيد. وكما يوحي المصطلح، فإن الأداة الموثوقة أو الثابتة هي الأداة التي تعمل بطرق متسقة ويمكن التنبؤ بها. ولكي يتمتع المقياس بالثبات، يجب أن تمثل الدرجات التي تنتج عن استخدامه بعضًا من الحالة الحقيقية للمتغير الذي يجري تقييمه وقياسه. ومن الناحية العملية، يعني هذا ضمناً أن الدرجة التي تم الحصول عليها من استخدام الأداة يجب ألا تتغير ما لم يكن هناك تغيير فعلي في المتغير الذي تقيسه هذه الأداة، وبالتالي، يمكن أن يعزى أي تغيير ملحوظ في الدرجات إلى التغيير الفعلي في هذا المتغير. ومن شأن وضع مقياس ثابت تمامًا أن يكون انعكاسًا للدرجة الحقيقية وليس أي شيء آخر. ونادراً ما يمكن تحقيق ذلك؛ ومع ذلك، يمكننا أن نقيس مدى قربنا من هذه الحالة المثالية. وكلما كانت الدرجة التي نحصل عليها من مقياس تمثل الدرجة الحقيقية للمتغير وكلما كانت تعكس عوامل خارجية أخرى بدرجة منخفضة، كان المقياس الذي لدينا أكثر ثباتاً. وبشكل أكثر رسمية، فإن ثبات المقياس هو نسبة التباين التي تُعزى إلى الدرجة الحقيقية للمتغير الكامن. وهناك العديد من الطرق والأساليب لحساب الثبات، ولكنها تشترك جميعاً في هذا التعريف الأساسي.

وعلى الرغم من أن الطرق البديلة لحساب الثبات قد تبدو مختلفة، فإن التعريف الأساسي المشترك يتطلب أن تكون متكافئة حسابياً بطريقه أساسيه وهامه. وهذا هو الحال بالفعل. وتتطوي كل هذه الأساليب والطرق على تقدير الدرجة الحقيقية للمتغير وتحديد نسبة درجة المقياس التي تم الحصول عليها والتي تمثلها الدرجة الحقيقية. ويقترح نموذج القياس الأساسي، والذي تم توضيحه في الفصل الثاني، أن الدرجة الملاحظة تمثل مجموع الدرجة الحقيقية للمتغير الذي يجري تقييمه

بالإضافة إلى الخطأ الناشئ عن عوامل خارجية. وتبعاً لذلك، نستطيع تقدير الدرجة الحقيقية للمتغير عن طريق طرح التباين الناتج عن الخطأ من التباين الكلي للدرجة الملاحظة التي تم الحصول عليها من مقياس معين. ويمكننا بعد ذلك حساب الثبات كنسبة للدرجة الحقيقية المقدرة إلى الدرجة الملاحظة. وهكذا:

$$\text{الدرجة الحقيقية (true score)} = \text{الدرجة الملاحظة (observed score)} - \text{الخطأ (error)}$$

$$\text{الثبات} = \text{الدرجة الحقيقية} / \text{الدرجة الملاحظة}$$

وتُعد طرق وأساليب تقدير الخطأ هي إلى حد كبير ما يميز المعادلات البديلة لحساب الثبات. وقد تم تصميم وتخصيص طرق مختلفة لأنواع محددة من البيانات، على الرغم من اشتراكها جميعاً في أساس مفاهيمي مشترك وهو: أن الثبات هو نسبة التباين في الدرجة الملاحظة التي يمكن أن تُعزى إلى الدرجة الحقيقية للمتغير الذي يجري تقييمه.

الطرق القائمة على تحليل التباين:

تعتمد إحدى وسائل تقدير الخطأ على تحليل التباين (ANOVA). ويقسم هذا المدخل التحليلي للبيانات التباين الكلي الملاحظ إلى مصادر مختلفة، لا سيما تلك التي تكون لها أهمية جوهرية (أي إشارة) وتلك التي تنشأ من مصدر خطأ ما (مثل الضوضاء)، مثل الخلل أو العيوب في أخذ عينة المشاركين من مجتمع ما. وعلى الرغم من أن هذا ليس هو المدخل الذي سنركز عليه في تقييم مدى ثبات المقاييس والقياس، فإن التطرق إليها بشكل موجز يؤكد على الاستمرارية عبر تعريفات ومداخل الثبات.

ومن ثم، من خلال استعراض ومراجعة سريعة، فإن النظر في مجموعة بسيطة للغاية من الملاحظات التي تنطوي على درجات حرارة ثمانية أجسام متماثلة، أربعة منها في ضوء الشمس المباشر وأربعة منها في الظل (لقد حددت عدداً صغيراً من الأجسام الملاحظة في هذا المثال من أجل التبسيط)، والأجسام متشابهة باستثناء تعرضها لأشعة الشمس؛ ومع ذلك، فإن مقياس الحرارة المستخدم لقياس درجة حرارتها مشكوك فيه بعض الشيء، وبالتالي، فهو مصدر محتمل للخطأ في درجات الحرارة الملاحظة. ويمكننا تقييم مدى هذا الخطأ من خلال تسجيل درجات حرارة كل جسم

من هذه الأجسام الثمانية على حدة وترتيب المعلومات بعدة طرق. أولاً، يمكننا تلخيص المعلومات حول الأجسام كمجموعة واحدة عن طريق حساب المجموع الكلي لمربع الانحرافات في درجات حرارة الجسم عن المتوسط الكلي لجميع الأجسام. ستكون هذه القيمة هي مجموع المربعات الكلي، أو SST. ومن خلال تقسيم مجموع المربعات الكلي على درجات الحرية المرتبطة بالعينة بأكملها (على سبيل المثال،  $N - 1 = 8 - 1 = 7$ )، سوف نحصل على التباين الكلي لدرجات حرارة الأجسام. والخطوات التالية تعزل المكونات الفرعية لهذا التباين الكلي. ويمكننا المضي قدماً في تقدير مدى تأثير الخطأ على هذه الدرجات وبالتالي المكون الفرعي من التباين الكلي. في إطار تحليل التباين ANOVA، يتم تحقيق ذلك من خلال تقييم مقدار التباين الذي يحدث في ظل ظروف مماثلة. في هذه الحالة، تتعرض جميع الأجسام الموجودة في الشمس لظروف متماثلة أو متطابقة، وكذلك جميع الأجسام الموجودة في الظل. داخل كل من هاتين المجموعتين الفرعيتين، يُفترض أن تكون الأجسام نفسها متطابقة وأن وجود أو عدم وجود ضوء الشمس متماثل. لذلك؛ يجب أن يكون الأساس الوحيد للاختلافات في درجات الحرارة الملاحظة هو نوع من أنواع الخطأ. وبالتالي، يمكننا دراسة وفحص التباين في درجات حرارة الأجسام داخل المجموعات لحساب مجموع مربعات الخطأ (SSE). وعن طريق طرح مجمع مربعات الخطأ SSE من مجموع المربعات الكلي SST، يمكننا حساب مجموع المربعات لتأثير ضوء الشمس. وهذا المجموع الأخير هو أساساً مجموع المربعات للدرجات الحقيقية، أي إنه مؤشر على مقدار التباين في درجات حرارة الجسم بعد إزالة تأثير خطأ القياس. ويمكننا بعد ذلك حساب تباين الدرجة الحقيقية من مجموع المربعات هذا. وأخيراً، من خلال حساب النسبة بين تباين الدرجات الحقيقية والتباين الكلي، نصل إلى نسبة التباين الكلي التي يمكن أن تعزى إلى الدرجة الحقيقية (أي تأثير الشمس). ويمكننا تفسير هذه النسبة على أنها ثبات قياسنا لدرجة حرارة الأجسام.

لاحظ أنه إذا كانت جميع الأجسام الموجودة في الشمس لها درجات حرارة متماثلة وكانت جميع الأجسام الموجودة في الظل لها درجات حرارة متماثلة، يفترض أنها منخفضة، فإن تباين الخطأ سيساوي القيمة “صفر”. وبالتالي، لن يتم طرح أي شيء من مجموع المربعات الكلي SST الملاحظة، وسيكون تباين الدرجة الحقيقية والتباين الكلي متساويين، وتكون النسبة التي تمثل ثبات قياس درجة حرارة الجسم تساوي الواحد الصحيح.

لقد أشرت إلى النسبة الناتجة عن مثال تحليل التباين الموضح أعلاه كمعامل ثبات، وهو صحيح. ومع ذلك، وبشكل أعم، فإن النسبة التي تقارن التباين الناتج عن بعض المصادر المحددة في تصميم تحليل التباين مع التباين الكلي تُعرف باسم معامل الثبات المعياري أو معامل الارتباط داخل الطبقة (intraclass correlation coefficient)، أو اختصاراً ICC. واعتماداً على نوع وتعقيد تصميم تحليل التباين، يمكن أن يكون هناك أنواع عديدة من معامل الارتباط داخل الطبقة والتي ستكون لها تفسيرات مختلفة، وليس كل منها تعادل أو تساوي ثبات القياس. على الرغم من أن القراء قد لا يكونون على دراية بمعامل الارتباط داخل الطبقة كما هو الحال بالنسبة للتعبيرات أو المصطلحات الأخرى الأكثر شيوعاً للثبات، إلا أننا سنرى أن المنطق الذي تستند إليه المزيد من المؤشرات الأكثر تخصصاً للثبات يتطابق مع منطق معامل الارتباط داخل الطبقة – وهذا يعني أن كلاً من معامل الارتباط داخل الطبقة والطرق الأخرى للحصول على الثبات تستند إلى مقارنة بعض تقديرات تباين الدرجة الحقيقية مع التباين الكلي.

#### البنود المتصلة «المستمرة» مقابل البنود ثنائية الاستجابة:

على الرغم من أن البنود قد تكون لها أشكال أو تنسيقات استجابة متنوعة، فإننا نفترض في هذا الفصل أن الاستجابات على البند تتألف من خيارات استجابة متعددة القيمة. وتستخدم البنود ثنائية الاستجابة (أي البنود التي لها خياران فقط للاستجابة، مثل «نعم» و«لا»، أو تلك التي لديها خيارات استجابة متعددة يمكن تصنيفها على أنها «صحيحة» مقابل «خاطئة») على نطاق واسع في اختبار القدرة، وبدرجه أقل، في سياقات قياس أخرى. وتتضمن أمثلة البنود ثنائية الاستجابة ما يلي:

1- زيوريخ هي عاصمه سويسرا. ( ) صحيح ( ) خطأ

2- ما هي قيمه «باي»  $\pi$ ؟

(أ) ١,٤١

(ب) ٣,١٤

(ج) ٢,٧٨

وقد تم تطوير أساليب وطرق خاصة لحساب الثبات والتي تستفيد من بساطة العمليات الحسابية للاستجابات ثنائية الخيارات. وتغطي كتب القياس العامة مثل “نونالي وبرنشتاين” (Nunnally and Bernstein, 1994) هذه الأساليب والطرق بشيء من التفصيل. إن منطق هذه الأساليب لتقييم الثبات يماثل إلى حد كبير المدخل الأكثر عمومية الذي ينطبق على البنود المتعددة الاستجابة والمقاييس المتصلة أو المستمرة. والواقع، إن المدخل المتبع في تقييم ثبات البنود متعددة الاستجابة هو في بعض الحالات امتداد لمدخل سابق تم وضعه وتطويره للبنود ذات الاستجابة الثنائية. وحرصاً على الإيجاز، فإن هذا الفصل لن يشير إلا بصورة عابرة إلى أساليب تقييم الثبات المعدة للقياسات التي تتكون من بنود ذات استجابات ثنائية. وتمت مناقشة بعض خصائص هذا النوع من المقاييس في الفصل الخامس.

#### الاتساق الداخلي:

إن ثبات الاتساق الداخلي Internal consistency reliability، كما يوحي الاسم، يتعلق بتجانس البنود داخل المقياس. وتهدف المقاييس القائمة على نماذج القياس الكلاسيكية إلى قياس ظاهرة واحدة. كما رأينا في الفصل السابق، تشير نظرية القياس إلى أن العلاقات بين البنود ترتبط منطقياً بعلاقات البنود بالمتغير الكامن. وإذا كانت لبنود مقياس ما علاقة قوية مع متغيرها الكامن، فستكون لها علاقة قوية مع بعضها البعض. وعلى الرغم من أننا لا نستطيع أن نلاحظ مباشرة العلاقة بين البنود والمتغير الكامن، إلا أنه يمكننا بالتأكيد تحديد ما إذا كانت البنود مرتبطة ببعضها البعض. ويكون المقياس متسقاً داخلياً بقدر ما تكون بنوده مرتبطة ارتباطاً عالياً ببعضها البعض. ما الذي يمكن أن يفسر الارتباط بين البنود؟ هناك احتمالان: إما أن تؤثر البنود سببياً على بعضها البعض (على سبيل المثال، يسبب البند “أ” حدوث البند “ب”)، أو أن البنود تشترك في قضية مشتركة. في معظم الحالات، يكون التفسير الأول غير مرجح، مما يجعل التفسير الأخير هو الخيار الأكثر وضوحاً. وهكذا، تشير الارتباطات العالية بين البنود وبعضها البعض إلى أن جميع البنود هي المقياس (أي، إنها مظاهر) لنفس الشيء. إذا وضعنا الافتراضات التي تمت مناقشتها في الفصل السابق (لا سيما الافتراض بأن البنود لا تتشارك في مصادر الخطأ)، يمكننا أيضاً استنتاج أن الارتباطات القوية بين البنود تنطوي على روابط قوية بين البنود والمتغير الكامن. ومن ثم، يجب أن يتكون المقياس الأحادي البعد أو بعد واحد لمقياس متعدد الأبعاد من مجموعة من البنود التي ترتبط جيداً مع بعضها البعض. المقاييس المتعددة الأبعاد التي تقيس عدة ظواهر - على سبيل

المثال، مقاييس موضع أو موقع التحكم أو السيطرة في الصحة (مقاييس موقع الضبط الصحي Health Locus of Control scales) (1) المتعددة الأبعاد (والستون، والستون، وديفيليس، Wallston, Wallston, & DeVellis, 1978) - هي في الواقع تجسد مجموعة من المقاييس المترابطة؛ وكل "بعد" هو مقياس بحد ذاته.

### معامل ألفا Alpha:

الاتساق الداخلي عادةً ما يتساوى (أو يعادل) مع معامل كرونباخ (Cronbach 1951) ألفا ( $\alpha$ ). وسوف ندرس معامل ألفا بشيء من التفصيل لعدة أسباب. أولاً، أنه يستخدم على نطاق واسع كمقياس للثبات. وثانياً، قد يكون ارتباطه بتعريف الثبات أقل وضوحاً مما هو الحال بالنسبة للمقاييس الأخرى للثبات (مثل أساليب النماذج البديلة) التي تمت مناقشتها مؤخراً. وبالتالي، قد يبدو معامل «ألفا» أكثر غموضاً من أساليب حساب الثبات الأخرى لأولئك الذين ليسوا على دراية بكيفية عمله الداخلي. وأخيراً، فإن استكشاف المنطق الذي يقوم عليه حساب معامل «ألفا» يوفر أساساً سليماً لمقارنه كيفية استيعاب الأساليب الحسابية الأخرى لجوهر ما نعنيه ونقصه بالثبات.

معادلة كودر - ريتشاردسون 20 (Kuder-Richardson formula 20) أو KR-20، كما هو معروف بشكل أكثر شيوعاً - هي نسخة خاصة من معامل «ألفا» للبنود ثنائية الاستجابة (انظر، علي سبيل المثال، نونالي وبرنشتاين، 1994 (Nunnally & Bernstein, 1994)). ويتساوى كل من معادلة كودر - ريتشاردسون KR-20 مع معامل «ألفا» عندما تكون البنود المكونة لمقياس هي بنود ثنائية الاستجابة. ومع ذلك، وكما ذكر آنفاً، سنركز على الشكل الأكثر عمومية الذي ينطبق على البنود التي لديها خيارات استجابة متعددة.

يمكنك التفكير في كل الاختلافات أو التباينات في مجموعه من درجات البنود على أنه يحدث بسبب أحد أمرين هما: (أ) التباين الفعلي بين الأفراد في الظاهرة التي يقيسها المقياس (أي التباين الحقيقي في المتغير الكامن) أو (ب) الخطأ. وهذا صحيح لأن نماذج القياس الكلاسيكية تعرف الظاهرة (على سبيل المثال، رغبة المرضى في التحكم أو السيطرة على التفاعلات مع الطبيب) كمصدر لكافة التباين المشترك، ويعرفون الخطأ على أنه أي تباين متبقي، أو غير مشترك، في درجات المقياس (على سبيل المثال، بند واحد يكون له معنى مزدوج غير المعنى المقصود). وهناك طريقة أخرى للتفكير في هذا الأمر وهو اعتبار أن التباين الكلي يشتمل على وجود اثنين من



المكونات هما: الإشارة (أي الفروق الحقيقية في رغبة المرضى في التحكم والسيطرة) والضوضاء (أي، الفروق في الدرجة الناتجة عن كل شيء ما عدا الفروق الحقيقية في الرغبة في التحكم أو السيطرة). إن حساب معامل «ألفا»، كما سنرى، يقسم التباين الكلي بين مجموعه البنود إلى مكونين هما الإشارة signal والضوضاء noise (الخطأ). نسبة التباين الكلي التي تمثل الإشارة تساوي معامل «ألفا». وهكذا، هناك طريقه أخرى للتفكير في معامل «ألفا» هو أنه يساوي الواحد الصحيح مطروحاً منه تباين الخطأ أو، على العكس من ذلك، فإن تباين الخطأ يساوي واحداً صحيحاً ناقص معامل ألفا (تباين الخطأ = 1 - معامل ألفا).

### مصفوفة التباين:

يساعدنا فحص مصفوفة التباين covariance matrix لمجموعة من بنود المقياس على تحقيق فهم الاتساق الداخلي بشكل كامل. وتكشف مصفوفة التباين لمجموعة من بنود المقياس عن معلومات مهمة حول المقياس ككل.

وتُعد مصفوفة التباين هي شكل أكثر عمومية لمصفوفة الارتباط correlation matrix. في مصفوفة الارتباط، تكون البيانات معيارية standardized (أي معايرة وحدات القياس)، مع تحديد التباينات بالواحد الصحيح. وفي مصفوفة التباين، تكون البيانات المدخلة غير معيارية (وحدات القياس الأصلية). وبالتالي، فإنه يحتوي على نفس المعلومات، كما في مصفوفة الارتباط، لكنها في شكل غير معياري، والبنود القطرية (المائلة) لمصفوفة التباين هي التباينات - تباينات البنود مع نفسها - تماماً مثلما تكون الوحدات الموضوعة على طول القطر الرئيسي لمصفوفة الارتباط هي تباينات المتغيرات المعيارية إلى الواحد الصحيح، وأيضاً ارتباطاتها مع نفسها. والقيم غير القطرية هي التباينات covariances، والتي تعبر عن العلاقات بين أزواج من المتغيرات غير المعيارية تماماً كما هو الحال في معاملات الارتباط المعيارية. وهكذا، من الناحية النظرية، تتكون مصفوفة التباين من (أ) التباينات variances (على القطر) لكل متغير من المتغيرات على حدة، و(ب) التباينات covariances (خارج القطر) التي تمثل العلاقة غير المعيارية بين أزواج المتغيرات.

وفي الجدول رقم 1-3 يتم عرض مصفوفة التباين النموذجية لثلاثة متغيرات هي  $X_1$  و  $X_2$  و  $X_3$ .

جدول (٣-1): التباينات والتغايرات لثلاثة متغيرات

$X_3$	$X_2$	$X_1$	
$Cov_{1,3}$	$Cov_{1,2}$	$Var_1$	$X_1$
$Cov_{2,3}$	$Var_2$	$Cov_{1,2}$	$X_2$
$Var_3$	$Cov_{2,3}$	$Cov_{1,3}$	$X_3$

ويتم عرض نفس المصفوفة بشكل أكثر توافقاً وتجانساً باستخدام الرموز المعتادة أو المألوفة للمصفوفات والتباينات والتغايرات على النحو التالي:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 \end{bmatrix}$$

مصفوفات التغير للمقاييس متعددة البنود:

دعونا نركز اهتمامنا على خصائص مصفوفة التغير لمجموعة من البنود التي تشكل، عند إضافتها معاً، مقياساً. تحتوي مصفوفة التغير الموضحة أعلاه على ثلاثة متغيرات هي  $X_1$ ، و  $X_2$ ، و  $X_3$ . افترض أن هذه المتغيرات هي في الواقع درجات لثلاثة بنود وأن هذه البنود هي ( $X_1$ ، و  $X_2$ ، و  $X_3$ ) عند إضافتها معاً تشكل مقياساً سنسميه ( $Y$ ). ما الذي يمكن أن نخبرنا به هذه المصفوفة عن علاقة كل بند من البنود على حدة بالمقياس ككل؟

تحتوي مصفوفات التغير على عدد من الخصائص المثيرة للاهتمام (على الأقل، الجيدة، والمفيدة). من بين هذه الخصائص حقيقة أن إضافة كل البنود في المصفوفات معاً (أي جمع التباينات، التي تقع على طول القطر للمصفوفات، والتغايرات، التي تقع خارج القطر) تعطي قيمة مساوية تماماً لتباين المقياس ككل، بافتراض أن البنود متساوية الأوزان. لذلك؛ إذا أضفنا كل البنود إلى مصفوفات التغير الرمزية، فسيكون المجموع الناتج هو تباين المقياس ( $Y$ ). هذا مهم للغاية وجدير بالذكر والتكرار: التباين في مقياس ( $Y$ ) والذي يتكون من أي عدد من البنود يساوي مجموع كافة القيم في مصفوفات التغير الخاصة بهذه البنود، بافتراض تساوي وزن البند ١.

وبالتالي، فإن تباين المقياس (Y) المكون من ثلاثة بنود متساوية في الوزن (X1 و X2 و X3) له العلاقة التالية بمصفوفات التباين الخاصة بالبنود، حيث:

$$C = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 \end{bmatrix}$$

القراء الذين يرغبون في الحصول على مزيد من المعلومات حول الموضوعات التي تم تناولها في هذا القسم يمكن إحالتهم إلى نونالي (Nunnally, 1978) من أجل مصفوفات التباين (التباين المشترك) ونامبوديري (Namboodiri, 1984) للحصول على مقدمة في مصفوفات الجبر في الإحصاء. تحتوي مصفوفات التباين لكل بند من البنود على حدة على معلومات مفيدة أخرى. وقد ناقشت "بوهرنستد" (Bohrnstedt, 1969) التطبيقات التي يمكن استخلاصها من مصفوفات التباين للبنود.

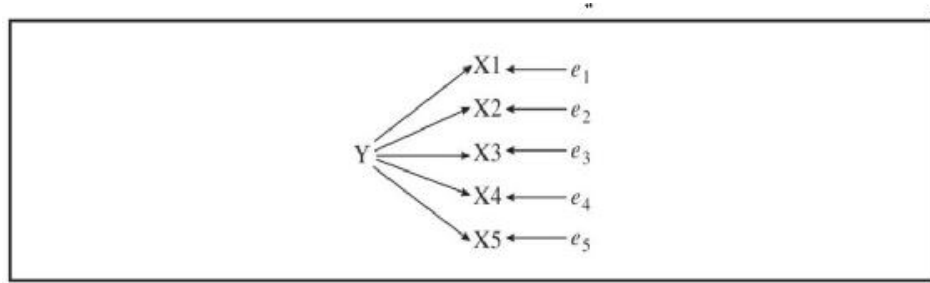
معامل «ألفا» ومصفوفة التباين:

تم تعريف معامل ألفا على أنه نسبة التباين الكلي للمقياس التي يمكن أن تعزى إلى مصدر مشترك، ومن المفترض أن تكون الدرجة الحقيقية لمتغير كامن تقوم عليه البنود. وبالتالي، إذا أردنا حساب معامل «ألفا»، فسيكون من المفيد الحصول على قيمة التباين الكلي للمقياس وقيمة لنسبة التباين «المشترك». مصفوفات التباين هي فقط ما نحتاج إليه من أجل القيام بذلك.

تذكر المخطط البياني الذي استخدمناه في الفصل الثاني لتوضيح كيفية ارتباط البنود بالمتغير الكامن الخاص بها، كما في الشكل رقم (٣-١).

جميع الاختلافات في البنود الناتجة عن المتغير الكامن (Y) هي مشتركة أو شائعة. (وتستخدم مصطلحات المشتركة joint والشائعة communal أيضاً لوصف هذا الاختلاف). عندما يختلف المتغير Y (كما هو الحال، على سبيل المثال، بين الأفراد الذين لديهم مستويات مختلفة من السمة التي يمثلها)، ستختلف الدرجات على كافة البنود معه لأنه سبب تلك الدرجات. لذلك، إذا كانت درجة المتغير (Y) مرتفعة، فسوف تكون كل درجات البنود مرتفعة؛ وإذا كانت درجة (Y)

منخفضة، فإن درجات البنود تميل إلى أن تكون منخفضة، وهذا يعني أن البنود سوف تميل إلى التغيير بشكل مشترك (أي، تكون مترابطة مع بعضها البعض). لذلك يؤثر المتغير الكامن على كافة البنود، وبالتالي فإن البنود تكون مرتبطة ببعضها البعض. وعلى النقيض من ذلك، فإن متغير الخطأ هو مصدر الاختلاف الفريد الذي يمتلكه كل بند. في حين أن كل البنود تشترك في التباين بسبب (Y)، فلا يوجد بندان اثنان يشتركان في أي تغيير من مصدر الخطأ نفسه في ظل افتراضات القياس الكلاسيكية الخاصة بنا. تؤثر قيمة متغير خطأ معين على درجة بند واحد فقط. وبالتالي، لا ترتبط متغيرات الخطأ ببعضها البعض. ولذلك؛ فإن كل بند (وضمنياً، المقياس الذي يحدده مجموع البنود) يتغير كدالة لـ (أ) مصدر التباين المشترك بينه وبين البنود الأخرى، و(ب) التباين الفريد وغير المشترك الذي نشير إليه على أنه خطأ. ويترتب على ذلك أن التباين الكلي لكل بند - وبالتالي، بالنسبة للمقياس ككل - يجب أن يكون مزيجاً من التباين من المصادر المشتركة والمصادر الفريدة. ووفقاً لتعريف الثبات، يجب أن تساوي قيمة معامل ألفا نسبة التباين في المصدر المشترك إلى التباين الكلي.



شكل (٣-١): مخطط يمثل كيفية ارتباط مجموعة مكونة من خمسة بنود بالمتغير الكامن المشترك (y)

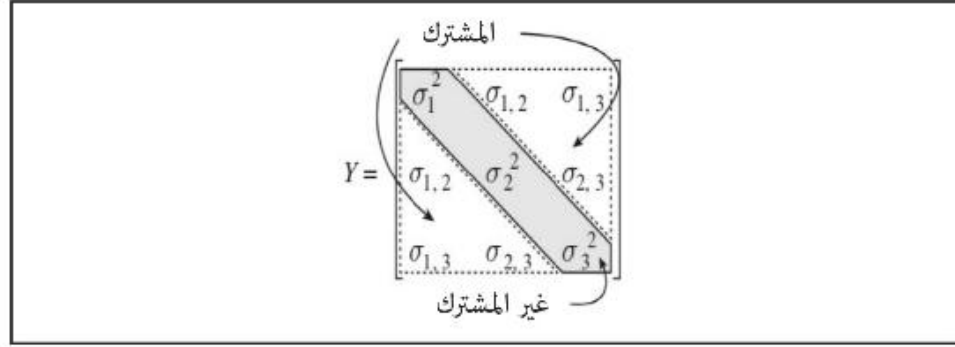
الآن، ضع في اعتبارك مقياساً من عدد (k) بند يسمى Y الذي تكون مصفوفة التباين له على النحو التالي:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \cdots & \sigma_{1,k} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{1,3} & \cdots & \sigma_{2,k} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \cdots & \sigma_{3,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,k} & \sigma_{2,k} & \sigma_{3,k} & \cdots & \sigma_k^2 \end{bmatrix}$$

التباين ( $\sigma^2$ ) للمقياس المكون من عدد ( $k$ ) بند يساوي مجموع جميع عناصر المصفوفة. المدخلات على طول القطر الرئيسي هي تباينات البنود الفردية الممثلة في المصفوفة. (ويُشار إلى تباين البند “i” كـ  $\sigma_i^2$ ). لذلك؛ فإن مجموع البنود على طول القطر الرئيسي ( $\sum \sigma_i^2$ ) هو مجموع التباينات لكل بند من البنود على حدة. وبالتالي، فإن مصفوفة التباين تتيح لنا وصولاً سهلاً إلى قيمتين هما: (1) التباين الكلي للمقياس ( $\sigma^2$ ) والذي يُعرف على أنه مجموع كل البنود في المصفوفة و(2) مجموع التباينات لكل بند من البنود على حدة ( $\sum \sigma_i^2$ ) المحسوبة من جمع القيم على طول القطر الرئيسي. وهاتان القيمتان يمكن إعطاؤهما تفسيراً مفاهيمياً أو نظرياً. فمجموع المصفوفة بأكملها، بحكم التعريف، هو تباين المتغير ( $Y$ )، للمقياس المكون من مجموعة بنود. ومع ذلك، فإن هذا التباين الكلي يمكن تقسيمه، كما قلنا، إلى أجزاء مختلفة.

دعنا ننظر في كيفية فصل مصفوفة التباين المشتركة عن التباين الفريد من خلال دراسة كيفية اختلاف البنود الموجودة في القطر الرئيسي لمصفوفة التباين عن كافة البنود غير القطرية (أي تقع خارج القطر). كل التباينات (البنود القطرية) هي متغير مفرد أو “متغير مع نفسه”. وقد لاحظنا في وقت سابق أن هذه التباينات يمكن اعتبارها بمثابة تباينات في البنود مع نفسها. ويحتوي كل تباين على معلومات حول بندٍ واحدٍ فقط. وبعبارة أخرى، يمثل كل منها معلومات تستند إلى بند واحد، وليس إلى التباين المرتبط المشترك بين البنود. (وفي إطار هذا البند الواحد، سيكون جزء من تباينه ناتجاً عن المتغير الأساسي المشترك، وبالتالي، سيتم تقاسمه مع بنود أخرى؛ وبعضها لن يكون كذلك. ومع ذلك، فإن تباين البند لا يحدد مقدار التباين المشترك ولكن فقط يحدد مقدار التشتت في الدرجات الخاصة بهذا البند، بغض النظر عن الأسباب). وتتضمن العناصر غير القطرية لمصفوفة التباين أزواجاً من البنود، وبالتالي، التباين الشائع (أو المشترك) بين بندين من بنود المقياس (التباين). وبالتالي، فإن العناصر الموجودة في مصفوفة التباين (وبالتالي التباين

الكلي للمتغير  $Y$ ) تتكون من التباين المشترك، إذا صح التعبير) بالإضافة إلى التباين “غير المرتبط” أو “غير المشترك” بشأن البنود التي يتم النظر فيها بشكل فردي. يمثل الشكل رقم (٣-٢) هذين التقسيمين الفرعيين لمصفوفة التباين. والمنطقة المظلمة على طول القطر هي الجزء غير المشترك من المصفوفة، والمنطقتان غير القطريتين داخل الحدود المثلثية هما، معاً، الجزء المشترك.



شكل (٣-٢) مصفوفة التباين - التباين تبين أن التباين على طول القطر الرئيسي

هو التباين غير المشترك، في حين أن التباينات الموجودة فوق القطر أو أسفله هي تباينات مشتركة

وبما أن التباينات - والتباينات فقط - تمثل التباين المشترك، فإنه يجب أن تكون كل التباينات غير المشتركة ممثلة في التباينات على امتداد القطر الرئيسي لمصفوفة التباين، وبالتالي، يمكن التعبير عنها بالمصطلح  $(\sum \sigma_i^2)$ . وبالطبع، يتم التعبير عن التباين الكلي بـ  $(\sigma_y^2)$ ، أي مجموع كافة عناصر المصفوفة. وبالتالي، يمكننا التعبير عن نسبة التباين غير المشترك إلى التباين الكلي في المتغير  $Y$ ) على النحو التالي:

$$\sum \sigma_i^2 / \sigma_y^2$$

وتتطابق هذه النسبة مع مجموع القيم القطرية في مصفوفة التباين. وبالتالي، يمكننا أن نعبر عن نسبة التباين المرتبط أو المشترك، وما يتبقى منها - وبعبارة أخرى، تكمة هذه القيمة كما هو مبين كالتالي:

$$1 - (\sum \sigma_i^2 / \sigma_y^2)$$

وتتوافق هذه القيمة مع مجموع كافة القيم غير القطرية لمصفوفة التباين. قد يبدو غريباً، أو على الأقل غير فعال، حساب العناصر القطرية ثم طرحها من قيمة مصفوفة التباين ككل. لماذا لا نحسب فقط مجموع العناصر غير القطرية بشكل مباشر مثل  $(\sum \sigma_{i,j})$ ، حيث إن (i) و (j) يمثلان كل بند من البندين المتضمنين في تباين معين؟ وفي الواقع، يمكن للمرء أن يصل إلى نفس النقطة بالضبط عن طريق الحساب المباشر لمجموع العناصر غير القطرية. إن الصيغة أو المعادلة التي تتضمن الطرح من الواحد الصحيح هي إرث من الأيام التي لم تكن فيها أجهزة الحاسب الآلي متوفرة لإجراء العمليات الحسابية.

حساب التباين الكلي للمتغير (Y) والتباين لكل عنصر على حدة (i) ربما كانت عمليات تم تنفيذها بالفعل لأغراض أخرى. وحتى لو لم تكن هناك حاجة لحساب هذه التباينات لأغراض أخرى، فانظر إلى الجهد الحسابي الذي تنطوي عليه. فبالنسبة للمقياس المكون من 20 بنداً، سيكون الاختيار بين حساب 21 تبايناً (واحد لكل بند والآخر للمقياس ككل) مقابل 190 تباين (أي واحد لكل عنصر من العناصر غير القطرية البالغ عددها 380 عنصراً في المصفوفة، مع تلك العناصر التي تكون فوق قطر المصفوفة المتماثلة لتلك التي أدناه) بالإضافة إلى التباين الكلي. ومن ثم، فإن الصيغة أو المعادلة التي تحدد التباين المشترك على أنه ما تبقى بعد إزالة التباين غير المشترك من التباين الكلي، تكون معادلة أكثر منطقية مما قد تبدو عليه في البداية.

القيمة التي تمثلها المعادلة هي:

$$1 - (\sum \sigma_i^2 / \sigma_y^2)$$

أو، أي ما يعادل أو يكافئ:

$$\sum \sigma_{i,j} / \sigma_y^2$$

من النظرة أو الوهلة الأولى قد يبدو أن ذلك يعكس تعريف معامل ألفا (أي الجزء المشترك من التباين الكلي في مقياس يمكن أن يعزى إلى المصدر المشترك للبنود، والذي نفترض أنه يعكس الدرجة الحقيقية للمتغير الكامن). بيد أننا ما زلنا بحاجة إلى تصحيح إضافي. وتصبح هذه الحاجة واضحة إذا أخذنا في الاعتبار ما سيحدث إذا كان لدينا، على سبيل المثال، خمسة بنود مترابطة

تمامًا. مثل هذا الترتيب ينبغي أن يؤدي إلى الثبات التام. إن مصفوفة الارتباط، كما ذكر آنفًا، هي مجرد مصفوفة تغاير حيث تم فيها معايرة «توحيد قياس» كل البنود. ونظرًا لأن البنود الموجودة في مصفوفة الارتباط أكثر ألفة واتساقًا (أي أن الارتباط التام يتطابق دائمًا مع معامل الارتباط الذي تكون قيمته الواحد الصحيح)، سأستخدم مصفوفة الارتباط لتوضيح هذه النقطة. إذا استخدمنا التعبيرات المذكورة أعلاه لتمثيل جزء من التباين المشترك والتباين الكلي في مصفوفة الارتباط الافتراضية الخاصة بها والمكونة من خمسة بنود لها ثبات تام، فما الذي يمكن الحصول عليه؟ تتكون مصفوفة الارتباط في هذه الحالة من مصفوفة من النوع  $(5 \times 5)$  وكل القيم تساوي الواحد الصحيح. ومقام المعادلة السابقة، الذي يمثل التباين الكلي للمقياس الذي يتألف من البنود الخمسة، يساوي 25. ومع ذلك، فإن البسط لا يساوي سوى 20 فقط، مما يسفر عن ثبات مقداره يساوي  $20/25$  (أو  $0.8$ ) بدلاً من الواحد الصحيح. لماذا يحدث هذا؟ العدد الإجمالي للعناصر في مصفوفة التغاير هو  $(2k)$ . وعدد العناصر في المصفوفة التي تكون غير مشتركة (على سبيل المثال، العناصر الموجودة على امتداد القطر الرئيسي) هو  $(k)$ . والرقم المشترك (كل العناصر غير القطرية) هو  $(-k^2)$ . وهكذا، فإن الكسر الموجود في المعادلة الأخيرة لدينا يحتوي على بسط يستند إلى قيم  $(-k^2)$  والمقام يستند إلى قيم  $(2k)$ . ولضبط أو تعديل حساباتنا بحيث تعبر النسبة عن الأهمية النسبية بدلاً من عدد البنود المجمعة في البسط والمقام، فإننا نقوم بضرب المعادلة بأكملها التي تمثل نسبة التباين المشترك في القيم المقابلة للفروق في عدد البنود التي تم تجميعها. وللقيام بذلك، نضرب المعادلة في المقدار  $(k^2) / (k^2 - k)$ ، أو، ما يساوي،  $(k - 1) / k$ . وهذا الإجراء يقيد مدى أو نطاق القيم المحتملة لمعامل ألفا لتكون ما بين الصفر والواحد الصحيح. في المثال المكون من خمسة بنود الذي تمت مناقشته للتو، يؤدي ضرب  $(0.8)$  في  $(5 / 4)$  إلى الحصول على القيمة واحد صحيح وهي قيمة الثبات التام. قد يرغب القراء في إجراء الحساب الذهني للمصفوفات ذات الأحجام الأخرى. ينبغي أن يصبح واضحًا أن المقدار  $(k - 1) / k$  هي دائمًا المضاعف (المضروب فيه) الذي سينتج عنه قيمة معامل ألفا مساوية لواحد صحيح عندما يكون الارتباط بين البنود وبعضها البعض تامًا. وهكذا، نصل إلى الصيغة أو المعادلة المعتادة لمعامل ألفا وهي:



$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum \sigma_i^2}{\sigma_y^2} \right)$$

وخلاصة القول، إن ثبات المقياس يساوي نسبة التباين الكلي بين بنوده التي تُعزى إلى المتغير الكامن وبالتالي فهي مشتركة. وتعتبر صيغة أو معادلة معامل ألفا عن ذلك لتحديد الجزء من التباين الكلي لمجموعة البنود الفريدة، مع طرح هذا من الواحد الصحيح لتحديد النسبة المشتركة، والضرب في معامل تصحيح لضبط وتعديل عدد العناصر المساهمة في الحسابات السابقة.

صيغة أو معادلة بديلة لمعامل «ألفا»:

هناك معادلة شائعة أخرى لحساب معامل ألفا تعتمد على الارتباطات بدلاً من التغيرات (التباينات المشتركة). في الواقع، إن هذه المعادلة تستخدم متوسط الارتباط بين البنود ( $\bar{r}$ ). هذه المعادلة هي:

$$\alpha = \frac{k\bar{r}}{1 + (k-1)\bar{r}}$$

وتتبع هذه المعادلة منطقياً للمعادلة القائمة على التغيرات لمعامل ألفا. تأمل أو خذ بعين الاعتبار معادلة التغيرات من حيث المصطلحات المفاهيمية:

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\text{مجموع تباينات البند}}{\text{مجموع التباينات والتغيرات}} \right)$$

لاحظ أن البسط والمقام في الحد على اليمين هما عبارة عن مجموع القيم لكل بند. ومع ذلك، فإن مجموع هذه القيم الفردية يكون مماثلاً لمتوسط القيم مضروباً في عدد القيم المعنية. (على سبيل المثال، أرقام (k) التي يساوي مجموعها 50 و (k) ضعف متوسط هذه الأرقام تساوي ٥٠ أيضاً. ولمزيد من التوضيح، بالتعويض عن k بالقيمة 10 في الإجراء السابق؛ يجب أن يكون متوسط 10 قيم يصل مجموعها إلى 50 هو خمسة أضعاف القيمة عشرة أو عشرة أضعاف القيمة خمسة، وهي نفس القيمة في المجموع الأصلي). لذلك؛ يجب أن يساوي البسط الخاص بالحد على اليمين (k)

مضروباً في متوسط تباين البند ( $\bar{v}$ ) ويجب أن يساوي المقام ( $k$ ) مضروباً في متوسط التباين زائد ( $k^2 - k$ ) - أو، بدلاً من ذلك، ( $k - 1$ ) ( $k$ ) - مضروباً في متوسط التباين ( $\bar{c}$ ):

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{k\bar{v}}{k\bar{v} + (k)(k-1)\bar{c}} \right)$$

لإزالة الواحد الصحيح من المعادلة، يمكننا استبدالها بما يعادلها،  $[k\bar{v} + (k)(k-1)\bar{c}] /$

$[k\bar{v} + (k)(k-1)\bar{c}]$  مما يسمح لنا بدمج الحد على اليمين بأكمله في نسبة واحدة:

$$\alpha = \frac{k}{k-1} \left( \frac{k\bar{v} + (k-1)\bar{c} - k\bar{v}}{k\bar{v} + (k)(k-1)\bar{c}} \right)$$

أي ما يعادل أو يكافئ:

$$\alpha = \frac{k}{k-1} \left( \frac{k(k-1)\bar{c}}{k[\bar{v} + (k-1)\bar{c}]} \right)$$

يؤدي الحذف المتبادل للمقدار ( $k$ ) من البسط الخاص بالحد الأيسر ومقام الحد الأيمن، بينما يؤدي الحذف المتبادل للمقدار ( $k - 1$ ) من البسط الخاص بالحد الأيمن ومقام الحد الأيسر إلى الحصول على الشكل المبسط التالي للمعادلة:

$$\alpha = \frac{k\bar{c}}{\bar{v} + (k-1)\bar{c}}$$

ونذكر أن الصيغة أو المعادلة التي نسعى للحصول عليها تنطوي على ارتباطات بدلاً من التباينات، وبالتالي متغيرات معيارية بدلاً من متغيرات غير معيارية. بعد المعايرة، يكون متوسط التباينات مطابقاً لمتوسط الارتباطات ويساوي التباين الواحد الصحيح. وبالتالي، يمكننا استبدال متوسط الارتباط بين البنود وبعضها البعض ( $\bar{r}$ ) مع الواحد الصحيح. وينتج عن ذلك الصيغة أو المعادلة القائمة على الارتباط لمعامل ألفا:

$$\alpha = \frac{k\bar{r}}{1 + (k - 1)\bar{r}}$$

وتُعرف هذه الصيغة أو المعادلة باسم معادلة تنبؤ سبيرمان - براون Spearman-Brown، وسيتم توضيح أحد استخداماتها المهمة في قسم من هذا الفصل الذي يتناول حساب الثبات بطريقة التجزئة النصفية. وتحسباً لهذه المناقشة، لاحظ أن المعادلة تقدم الثبات كدالة لخاصيتين من خصائص المقياس هما: عدد البنود في المقياس (k) ومتوسط الارتباط بين أزواج البنود ( $\bar{r}$ ). وبالتالي، كلما زاد عدد البنود التي يحتويها المقياس وكان متوسط الارتباط بين تلك البنود أقوى، زاد الثبات.

ويشار أحياناً إلى الصيغتين المختلفتين، التي تستند إحداهما على التغيرات وتستند الأخرى على الارتباطات، على أنهما معادلات وصيغ الدرجة الخام والدرجة المعيارية لمعامل ألفا، على الترتيب. وتحافظ معادلة الدرجة الخام على المعلومات حول متوسطات وتباينات البند في عملية الحساب لأن التباينات المشتركة (التغاير) تستند إلى القيم التي تحتفظ بوحدات القياس الأصلية للبيانات الخام. إذا كانت للبنود تباينات مختلفة بشكل ملحوظ، فإن تلك البنود التي لها تباينات أكبر سيكون لها وزن أكبر من تلك البنود التي لها تباينات أقل عند استخدام هذه المعادلة لحساب معامل ألفا. ولا تحتفظ المعادلة المعيارية أو القياسية القائمة على الارتباطات بوحدات القياس الأصلية للبنود. تذكر أن الارتباط هو تغاير معياري أو قياسي. لذلك يتم وضع جميع البنود على مقياس مشترك، وبالتالي، تكون الأوزان متساوية في حساب معامل ألفا من خلال المعادلة المعيارية. أيهما أفضل أن يتم الاعتماد على السياق المحدد أم الوزن المتساوي؟ وكما سنرى في الفصول اللاحقة، فإن الإجراءات الموصى بها لتطوير البنود غالباً ما تستلزم هيكلة صياغتها بحيث تسفر عن تباينات مماثلة لكل بند. وعندما يتم اتباع هذه الإجراءات، يكون هناك عادةً اختلاف بسيط في معاملات ألفا المحسوبة بواسطة الطريقتين البديلتين. ومن ناحية أخرى، عندما لا يتم اتباع الإجراءات الرامية إلى إحداث تباينات متكافئة في البنود، يلاحظ أن قيم ألفا المعيارية والخام تختلف اختلافاً كبيراً (على سبيل المثال، (0.05) أو أكثر) يدل على وجود بند واحد على الأقل له تباين يختلف اختلافاً كبيراً عن التباينات في البنود الأخرى.

نقد معامل «ألفا»:

يمكن للمرء أن يجادل بأن معامل ألفاكرونباخ ليس مؤشراً مثالياً للثبات (على سبيل المثال، سيجتزما 2009, Sijtsma). وفي الواقع، فإنه - أي معامل ألفاكرونباخ - عادةً ما يوفر الحد الأدنى لتقدير الثبات lower bound الحقيقي لمجموعة من البنود بدلاً من أفضل تقدير للثبات الحقيقي. ويشير «سيجتزما» (٢٠٠٩) إلى أن مفهوم الاتساق الداخلي غامض إلى حد ما ويثير تساؤلاً حول مدى تمثيل معامل ألفا لأهم جوانب الاتساق الداخلي. ويشير إلى أن الاتساق الداخلي يتعلق بالبنية العاملية لمجموعه من البنود وأن اعتبار معامل ألفا مؤشراً على ذلك هو إفراط في التبسيط. بالمعني الدقيق أو الضيق، فإن رأي «سيجتزما» صحيح. غير أنه يعترف بوجود مشاكل في التقديرات الأخرى (على سبيل المثال، تقدير الحد الأدنى المنخفض للثبات، greatest-lower-bound estimate of reliability, or glb) التي يمكن اعتبارها بديلاً عن معامل ألفا. ويقترح «سيجتزما»، بشكل مثير ومناسب، كسبيل إلى المزيد من الدقة، ينبغي القيام بمزيد من العمل من أجل الفهم الكامل لأنواع التحيز التي قد تكون موجودة في مختلف مؤشرات الثبات.

هناك نقد آخر لمعامل ألفا كمعامل للثبات يتمثل في أنه تم تصميمه بشكل أساسي للبيانات المستمرة أو المتصلة. وتعتمد الحسابات التي تنطوي عليها عملية تحديد معامل ألفا على ارتباطات بيرسون Pearson، التي تفترض القياس المسافي أو الفتري scaling interval. ويلاحظ كل من “جادرمان، وجوهن، وديمبو” (Zumbo and, Guhn, Gadermann, 2012) أن أنواع الاستجابات أو الردود على البنود كثيراً ما تُستخدم في العلوم الاجتماعية وفي غيرها هي، على وجه الدقة، من نوع القياس الترتيبي وليست من نوع القياس المسافي أو الفتري. وكمثال، خذ بعين الاعتبار البند الافتراضي التالي (وخيارات الاستجابة المصاحبة له) والذي يقصد به أن يكون جزءاً من مقياس الشك الذاتي:

لم أتمكن من إنجاز وتحقيق كل ما كنت أتمنى أن أفعله.

٦	٥	٤	٣	٢	١
أوافق بشدة	أوافق بدرجة متوسطة	أوافق بدرجة قليلة	لا أوافق بدرجة قليلة	لا أوافق بدرجة متوسطة	لا أوافق بشدة

من الواضح أن الفئات مرتبة من الأقل اتفاقاً إلى الأكثر اتفاقاً مع البند، ولكننا لا نستطيع أن نفترض أن الزيادة في الشك الذاتي التي يمثلها الفرق بين اختيار الاستجابة (٣) بدلاً من الاستجابة (٢) هو نفس مقدار الزيادة في الشك الذاتي عند اختيار الاستجابة (٦) بدلاً من الاستجابة (٥). أي، إنه لا يمكننا أن نفترض زيادات المسافات بالتساوي في المتغير الذي يستتبطه البند. فالفئات مجرد فئات ترتيبية بالمعنى الدقيق للكلمة. وقد أشار «جادرمان وآخرون» (Gadermann et al., 2012) إلى أن هذه المسافات غير المتساوية تؤدي إلى عدم الدقة في استخدام معامل ألفا كتقدير لثبات الاتساق الداخلي وأوصوا ببديل آخر، أطلقوا عليه معامل ألفا الترتيبي ordinal alpha coefficient، والذي لا يفترض القياس الفتري أو المسافي. ويستبدل مدخلهم أو طريقتهم هذه استخدام ارتباطات بيرسون بالارتباطات الرباعية tetrachoric أو متعددة الأقسام polychoric (الأولى في الحالات التي تنطوي على الاستجابات الثنائية للبند، والأخيرة في الحالات التي تنطوي على خيارات استجابة ترتيبية متعددة النقاط). وتفترض الارتباطات الرباعية / متعددة الأقسام أن المتغير قيد الدراسة هو متغير مستمر أو متصل لكن يتم قياسه بشكل ترتيبي، باستخدام فئتين أو أكثر مرتبة كخيارات استجابة. ومن الناحية النظرية، يمكن للمرء أن يفكر في حدود السمة المقاسة (الشك الذاتي في هذا المثال) الموجودة على طول سلسلة متصلة. عند مستوى ما من السمة، هناك حد، عندما يتم تجاوزه، يميل المجيب لاختيار خيار الاستجابة (٢) بدلاً من الخيار (1). توجد حدود إضافية (عند مسافات غير محددة) بين كل خيار استجابة والبديل الأعلى التالي. إن الارتباط متعدد الأقسام بين أي بندين من النوع الذي تم وصفه سابقاً يبين أساساً مدى ميل هذه الانتقالات والتحويلات عبر هذه الحدود إلى الحدوث بشكل مرتبط بدلاً من أن تحدث بشكل مستقل بالنسبة للبنود المعنية. هذا الارتباط هو تقدير الاحتمال الأقصى للارتباط الحقيقي بين المتغيرات الترتيبية (تيمرمان، ولورينزو و Lorenzo-Seva, & Timmermann, 2011). باستبدال مصفوفة ارتباطات بيرسون بمصفوفة الارتباطات متعددة الأقسام، يمكن للمرء الحصول على معامل ألفا الترتيبي بدلاً من معامل ألفا المعياري.

وقام «غادرمان وآخرون» (Gadermann et al., 2012) ببحث تأثير استبدال ارتباطات بيرسون بالارتباطات الرباعية / متعددة الأقسام في دراسة تتضمن بيانات محاكاة data simulated. أظهرت نتائج دراسة المحاكاة هذه أن معامل ألفا يظهر تحيزاً أكبر (أي انخفاض التقديرات لأقصى درجة للثبات المعروف لمجموعة البنود المعنية) مقارنةً بمعامل ألفا

الترتيبي. إن الالتواء الشديد في توزيع الدرجات يضر على وجه الخصوص بدقة التقديرات الناتجة عن معامل ألفا كرونباخ، وفقاً للمحاكاة التي أوردها «غادرمان وآخرون»، ومع ذلك، تجدر الإشارة إلى أن الارتباطات متعددة الأقسام تنطوي على قيود خاصة بها، لا سيما مع أحجام العينات الصغيرة (Timmermann & Lorenzo-Seva, 2011).

انتقد مؤلفون آخرون معامل ألفا لأسباب أخرى غير تحيزها عندما يكون قياس البند من النوع الترتيبي وليس من النوع الفكري أو المسافي. على سبيل المثال، لاحظ كل من «ديون، باجولي، وبرينسدن» (Dunn, Baguley, and Brunsden, 2014: p. 404) ما يلي:

الصعوبات الرئيسية في استخدام معامل ألفا كمقياس للاتساق الداخلي يمكن تلخيصها على النحو التالي:

- ١- يعتمد معامل ألفا على افتراضات بالكاد يمكن أن تتحقق (أي يصعب تحقيقها).
- ٢- يؤدي انتهاك هذه الافتراضات إلى تضخيم معامل ألفا وتقليل تقديرات الاتساق الداخلي للمقياس.
- ٣- «ألفا إذا تم حذف بند» [طريقة عامة لتقييم تأثير كل بند من البنود على حدة على معامل ألفا الكلي] في عينة لا يعكس تأثير حذف البند على ثبات المجتمع الإحصائي population.
- ٤- لا يعكس التقدير بنقطة لمعامل ألفا التباين الموجود في عملية التقدير، مما يوفر ثقة زائفة في اتساق إدارة وتطبيق مقياس ما.

سأتناول كلاً من هذه الانتقادات تباعاً واحدة بعد الأخرى. تتطلب افتراضات نموذج «تاو» المتماثل أو المترادف في الأساس التي يقوم عليها معامل ألفا أن تكون كل البنود مؤشرات جيدة بنفس القدر لمفهوم بنائي واحد يشتركون فيه. أي، إنه يجب أن يقوم أي بند بمهمة جيدة في التأثير على السمة التي يتم قياسها مثل أي بند آخر. وبشكل أكثر رسمية، يحدد هذا الافتراض أن مجموعة البنود تكون أحادية البعد وأن التغيرات بين البنود والدرجة الحقيقية للمتغير الكامن متساويان أساساً، على الرغم من أن تباينات الخطأ (وبالتالي التباينات الملاحظة للبنود) لا يلزم أن تكون كذلك. ويمكن أن يحدث انتهاك أو مخالفة الافتراض الخاص بأحادية البعد إذا كان محتوى البنود

ينطوي على أكثر من مفهوم واحد متميز تشترك فيه البنود. على سبيل المثال، قد يؤدي استخدام أداة تهدف إلى قياس القلق إلى قياس جوانب الاكتئاب عن غير قصد. وإذا لم يتم إدراك وتحديد هذه الازدواجية المفاهيمية، فإن ذلك قد يشكل انتهاكًا لافتراض أحادية البعد الذي يقوم عليه استخدام معامل ألفا. وبطبيعة الحال، فإن افتراض أحادية البعد هو أيضًا افتراض لمعظم البدائل المقترحة لمعامل ألفا، وبالتالي، فإن انتهاك افتراض أحادية البعد يطرح مشكلة أكثر عمومية في تقدير الثبات. وهناك نوع أكثر شيوعًا من انتهاك نموذج «تاو» المتمثل أو المترادف في الأساس هو وجود بنود لا تعبر عن المتغير الأساسي بنفس القدر من الصحة، أي أن لها تغيرات مختلفة مع الدرجة الحقيقية. وسيتضح ذلك في شكل ارتباطات متبادلة غير متجانسة بين البنود. لأن كل التباين المشترك بين بندين في مجموعة بنود أحادية البعد، هو مظهر من مظاهر ارتباط كل بند بالدرجة الحقيقية، وإذا كانت بعض البنود ارتباطها أقل قوة بتلك الدرجة الحقيقية من غيرها، فإن الارتباطات بين البنود التي تنطوي على تلك البنود ستكون أقل. لذلك؛ بالنسبة لأربعة عناصر افتراضية هي A و B و C و D لها تغيرات خاصة بكل منها مع الدرجات الحقيقية هي ٠,٣١ و ٠,٣٠ و ٠,٢٧ و ٠,٠٥، فإن الارتباطات بين A و B و C (أي  $r_{AB}$  و  $r_{AC}$  و  $r_{BC}$ ) بشكل عام ستكون أعلى (وأكثر تشابهًا) من الارتباطات التي تتضمن البند D (أي  $r_{AD}$  و  $r_{BD}$  و  $r_{CD}$ ). سيكون عدم تجانس تباين البنود بمثابة دليل مهم على انتهاك افتراض نموذج «تاو» المتمثل أو المترادف في الأساس (أي، مقارنات التغيرات من البند إلى الدرجة الحقيقية). كما أن البنود التي لها خيارات استجابة مختلفة، مثل خيار مكون من (٤) نقاط لبعض البنود وخيار مكون من (٧) نقاط للبنود الأخرى، كثيرًا ما تؤدي أيضًا إلى انتهاك افتراض تماثل تغيرات البند والدرجة الحقيقية. لماذا هو كذلك؟ وحتى عندما يكون هناك شخصان لهما نفس الدرجة الدقيقة من السمة التي يجري تقييمها، فإن التطابق بين درجاتها على بندين لا يمكن أن تكون تامة إلا عندما يكون عدد خيارات الاستجابة متطابقًا لكلا البندين. وبشكل عام، فإن عدم تطابق عدد خيارات الاستجابة بين بندين سيسفر عن انخفاض درجة ارتباطهما معًا، مقارنةً بأداء نفس البنود عندما يكون لكل منهما نفس العدد من خيارات الاستجابة. لذلك؛ ستؤدي خيارات الاستجابة غير المتطابقة إلى بعض أزواج البنود (تلك التي لها نفس عدد خيارات الاستجابة) التي من المحتمل أن يكون لها ارتباطات أقوى مع أزواج بنود أخرى (حيث لا يتطابق عدد خيارات الاستجابة). وبما أن الارتباطات بين البنود وبعضها البعض متصلة في ارتباطات البنود بالدرجة الحقيقية، فإن هذه الارتباطات المتباينة بين البنود

وبعضها البعض تعني ضمناً أن التغيرات بين البنود والدرجة الحقيقة تكون مختلفة أيضاً عبر البنود. هذا، بطبيعة الحال، يشكل انتهاكاً لنموذج «تاو» المتماثل أو المترادف في الأساس والذي يستند إليه معامل ألفا. ونتيجة لانتهاك هذا الافتراض المحدد، فإن معامل ألفا قد لا يقدم تقديراً دقيقاً للثبات الحقيقي للمقياس. وهكذا، يجادل دان وآخرون (Dunn et al., 2014) بأن استمرار استخدام ألفا يطرح (أو يسبب) مشكلة.

أما حجتهم الثانية هي أن الفشل في تلبية تلك الافتراضات والوفاء بها ليست مجرد مسألة نظرية بل هي أيضاً نتيجة عملية. أي إن أنواع الثبات تقدر التباينات التي ينتجها معامل ألفا كبيرة بما فيه الكفاية. على الرغم من أن كلاً من المغالاة في التقديرات والتقليل من التقديرات هما أمران ممكنان عندما تفشل مجموعه البنود في الوفاء بافتراضات نموذج «تاو» المتماثل أو المترادف في الأساس، فإن التقدير المنخفض هو الأكثر شيوعاً بكثير، لأن معامل ألفا هو تقدير الحد الأدنى للثبات في إطار نموذج «تاو» المتماثل أو المترادف في الأساس. ومع ذلك، يمكن أن يحدث التضخم عندما يكون عدد البنود التي تشكل المقياس كبيراً جداً (لأن القيمة التي تم الحصول عليها لمعامل ألفا تميل إلى الزيادة مع عدد البنود). أما الظروف التي تنطوي على أكبر التناقضات بين مستويات ألفا الاسمية والثبات الحقيقي هي تلك التي تشكل انتهاكات صارخة للافتراضات المتعلقة لنموذج «تاو» المتماثل أو المترادف في الأساس. وقد تم وصف بعض هذه الحالات في وقت سابق عند «غادرمان وآخرين» (Gadermann et al., 2012).

أيضاً، انتقد «دان» وآخرون (Dunn et al., 2014) الممارسة الخاصة بحذف البنود بالتتابع واحداً تلو الآخر، ثم النظر في التأثير على معامل ألفا كوسيلة لاكتشاف أفضل السبل لتحسين الاختصار للمقياس وثباته (موضوع سأناقشه بمزيد من التفصيل في فصل لاحق). وتؤكد الحجة التي طرحوها على أن حذف بند معين قد يزيد معامل ألفا لمجموعة متنوعة من الأسباب، وبالتالي قد لا تمثل زيادة في تغاير الدرجة الحقيقية للبند ككل. ويقترحون أن الأساس المنطقي لحذف بند ما يستند على افتراض أن جميع البنود لها تباينات خطأ متساوية (وهي خاصية أو سمة من سمات البنود التي تفي أو تلبي افتراضات الاختبارات المتوازية أو المتكافئة ولكن ليست أساساً للاختبارات الخاصة لنموذج «تاو» المتماثل أو المترادف في الأساس، على النحو المشار إليه في الفصل الثاني). ولكن هذا لا يفترض في معظم تطبيقات معامل ألفا. وغالباً، يكون الافتراض هو لنموذج «تاو» المتماثل أو المترادف في الأساس، والذي لا يتطلب تباينات خطأ متساوية. وبالتالي، قد لا يكون



للبيد المحذوف تباين درجة حقيقية أكبر، ولكن يملك تباين خطأ أقل فقط عند استخدامه في هذه الحالة المحددة (أي، مع العينة التي تم جمع البيانات منها). ومن المفترض أن التقلبات أو التغيرات في تباين الخطأ تعتمد بشكل أكبر على العينة، في حين أن التباين في الدرجة الحقيقية أكثر صلة بتقدير مدى ثبات المقياس في المجتمع الإحصائي. ومن ثم، فإن الحجة، تذهب إلى القول بأن التحسن الظاهر في الثبات الذي تحقق عن طريق إسقاط أو حذف البنود التي لا تعزز معامل ألفا قد يكون مجرد وهم أو خداع.

ويستند انتقادهم النهائي لمعامل ألفا على فكرة أن تقدير قيمة وحيدة للثبات ليست ذا مغزى محدد ما لم يكن لدينا بعض الفهم والإدراك بمدى التششت الكبير لمثل هذه التقديرات. أي إنه بدون فترة ثقة حول تقدير الثبات الذي يوفره معامل ألفا، فإنه يكون من الصعب قياس مقدار الثقة التي ينبغي أن تكون لدينا في ذلك التقدير.

#### معالجة محددات وقيود معامل «ألفا»:

إحدى وسائل التغلب على هذا الانتقاد الأخير هي تحديد فترة ثقة لمعامل ألفا. وهذا يسمح للباحث بأن يتأكد أنه ضمن هامش خطأ محدد مسبقاً (على سبيل المثال، الثقة بنسبة 95%)، يقع الثبات الحقيقي للمقياس بين قيمتين محددتين. لذلك؛ على سبيل المثال، إذا كانت قيمة معامل ألفا تبلغ ٠,٨٥ (تقدير النقطة للثبات) لها فترة ثقة 95 ٪ يحدها ٠,٨٠ و ٠,٩٠، فإن ذلك يشير إلى أن احتمال أن يقع الثبات الحقيقي داخل تلك الحدود كان بنسبة 95 ٪ واحتمال أن يكون خارج تلك الحدود كان بنسبة 5%.

الطريقة التي أوصى بها «دان وآخرون» (Dunn et al., 2014) وغيرهم (على سبيل المثال، باديل، وديفرز، ونيوتون Padilla, Divers, & Newton, 2013) لتحديد فترات الثقة لتقدير الثبات المعين الذي يتضمن تقنية <sup>3</sup> bootstrapping. ولتوضيح تقنية بوتستراينج bootstrapping، أقوم أولاً بمراجعة كيفية إنشاء فترة ثقة بدون بوتستراينج. وعادةً ما يتطلب مفهوم فترة الثقة، كما ينطبق هنا، استخدام مجموعات بيانات متعددة من أجل الحصول على توزيع قيم لمعلمة المجتمع parameter التي يتم تقديرها. لذلك؛ على سبيل المثال، إذا أردنا فترة ثقة حول تقدير نقطة لمتوسط طول للأشخاص في مدينة معينة، فيمكننا سحب أو أخذ العديد من عينات الأشخاص من تلك المدينة، وحساب متوسط الطول لكل عينة، وإنشاء توزيع لهذه المتوسطات.

ويكون المتوسط الكبير (أي القيمة المتوسطة لتوزيع متوسطات العينة) هو تقديرنا لنقطة المتوسط الحقيقي لطول الأشخاص في المدينة. ونتوقع حدوث اختلافات في متوسط الأطوال الملاحظة من عينة إلى أخرى بشكل عشوائي لأن جميع العينات مأخوذة من نفس المجتمع. وبالتالي، ينبغي أن يكون توزيع المتوسطات لمتوسط الطول توزيعاً طبيعياً (خاصةً إذا قمنا بسحب أو أخذ عدد كاف من العينات). وكما هو الحال مع أي توزيعات طبيعية، يمكننا تحديد الحدود التي كانت بالنسبة لـ 95٪ المركزية من متوسط الأطوال، مع تحديد قيم العينات المقابلة للحدود الدنيا والعليا لهذه المنطقة. لذلك؛ إذا كان المتوسط الكلي (على سبيل المثال، متوسط كل متوسطات العينات الفردية) هو 68 بوصة، فيمكننا إيجاد الحدود العليا والدنيا لفترة الثقة البالغة 95٪ حول تقدير هذه النقطة. وتتبع بوتستراينج Bootstrapping نفس المنطق مع اختلاف مهم للغاية. فبدلاً من أخذ عينات متعددة من المجتمع، فإنه يحاكي هذه العملية من خلال أخذ عينات فرعية متعددة من عينة واحدة متوفرة. وفي الأساس، يتم التعامل مع العينة التي تكون في متناول اليد كمجتمع ويتم السحب مع الإرجاع replacement لعدد كبير من "عينات" مختلفة منها. بهذه الطريقة، يتم إنشاء إصدارات أو نسخ مختلفة من العينة ويمكن حساب معامل ألفا في كل منها، مما يؤدي إلى توزيع درجات معامل ألفا التي يمكن أن تكون بمثابة أساس لتحديد فترات الثقة.

وتقدم باديللا وآخرون (Padilla et al., 2013) وصفاً مفصلاً لطريقة بوتستراينج لحساب فترات الثقة لمعامل ألفا باستخدام بيانات المحاكاة. وكوسيلة لفهم فائدة طريقة بوتستراينج، يمكن أن تكون بيانات المحاكاة مفيدة بشكل خاص لأن مجموعة البيانات يمكن إنشاؤها مع الثبات «الحقيقي» للمقياس الافتراضي المحدد مسبقاً من قبل الباحثين. ويمكن تحديد قيمة درجة حقيقية تعسفية كما يمكن توليد بيانات «ملاحظة» تختلف عن الدرجة الحقيقية بطرق تخضع تماماً لسيطرة أو ضبط الباحثين. وبالتالي، يمكن للباحثين التحكم في كل من مدى ثبات المقياس الافتراضي من البداية وخصائص التوزيع (مثل الالتواء والتفلطح) لمجموعة بيانات المحاكاة. وبعد ذلك، يمكن للباحثين تنفيذ طرق مختلفة للبوتستراينج وتحديد أيهم يوفر تمثيلاً أكثر دقة للقيم «الحقيقية» التي قاموا ببنائها من بيانات المحاكاة. واستخدمت «باديللا وآخرون» (2013) هذا المدخل لمقارنة طرق مختلفة لتقدير فترات الثقة لمعامل ألفا تحت ظروف بيانات مختلفة (على سبيل المثال، بيانات ذات توزيع غير طبيعي). وتشير نتائجها إلى أن بعض الطرق تعمل بشكل جيد عبر مجموعة متنوعة من الظروف، في حين أن بعض الطرق الأخرى تعتمد بدرجة أكبر على شروط توزيعية معينة للتأكد

من دقتها. وبشكل أعم، فإن النتائج التي توصلوا إليها، استناداً إلى عمليات المحاكاة، توفر رؤى تفصيلية عن الوقت الذي يمكن فيه تطبيق مدخل البوتستراينج لإيجاد فترات الثقة لمعامل ألفا كتقدير لمدى الثبات.

تتمثل أحد الجوانب السلبية المحتملة لاستخدام المحاكاة في أنه من السهل وغالباً ما يكون من المغري محاكاة خصائص التوزيع المتطرفة التي تنتج اختلافات كبيرة بين المعلمات المقدرة والمعلمات النظرية مثل الثبات. على الرغم من أن أهمية أسلوب المحاكاة، إلا أن أكثر ما يهتم به الباحثون هو خصائص بياناتهم الفعلية. لحسن الحظ، يمكن أيضاً استخدام مدخل البوتستراينج bootstrapping مع البيانات الفعلية لتحديد فترات الثقة حول قيمة معامل ألفا التي تم الحصول عليها. لم يتم استخدام Bootstrapping على نطاق واسع بهذه الطريقة في الماضي، ويرجع ذلك جزئياً إلى أن الأدوات الحسابية لإنشاء نماذج فرعية من bootstrap ولحساب فترات الثقة لم تكن متوفرة بسهولة. ولقد تغير هذا إلى حد ما مع ظهور برنامج التحليل الإحصائي (Venables, R Smith, & the R-Core team, 2015) وهي بيئة برمجية قوية ومتعددة الاستخدامات تمكن المستخدمين من أداء مجموعة واسعة من عمليات التحليل والرسومات البيانية. وبرنامج التحليل الإحصائي R مجاني ويمكن الوصول إليه من خلال الرابط التالي: <http://www.r-project.org>. هذا الموقع يوفر أيضاً الوثائق. وداخل بيئة البرنامج الإحصائي R، يوجد عدد من الحزم المتاحة والتي هي في الأساس مجموعات من البرامج الفرعية المصممة حول تحقيق أهداف تحليلية أو رسومية محددة. وهذه أيضاً متاحة مجاًناً ويمكن الوصول إليها من مواقع نسخ متطابقة متعددة من داخل بيئة البرنامج الإحصائي R.

#### معامل أوميغا ( $\omega$ ):

على الرغم من أنه من الواضح إمكانية تعزيز فائدة معامل ألفا باستخدام الارتباطات الرباعية /المتعددة الأقسام أو عن طريق تحديد فترات الثقة، يوصي العديد من المتخصصين في منهجية البحوث بمدخل بديل يعالج محددات أو قيود معامل ألفا بشكل كامل. هذا العلاج أو الوسيلة تتمثل في معامل أوميغا ( $\omega$ )، وهو تقدير بديل للثبات. ومنطق معامل أوميغا مشابه لمنطق معامل ألفا. وكلاهما يعرفان الثبات أساساً على أنه نسبة تباين الدرجة الحقيقية، مقارنة بالتباين الكلي الملاحظ. ومع ذلك، يستخدم معامل أوميغا طريقه مختلفة لحساب التباين ذي الصلة. في حين يشتق

معامل ألفا تقديرات التباين من مصفوفة التباين (أو الارتباط) للبنود التي تشكل مقياساً، فإن معامل أوميغا يستخدم مصفوفة من تشعبات البنود على عامل مشترك وحيد تشترك فيه البنود. وستتم مناقشة التحليل العاملي بشكل أكثر شمولاً في فصل لاحق، ولكن في الوقت الراهن، يمكننا أن نؤكد أن أحد استخداماته هو تقدير الارتباط بين كل مجموعه من البنود، من ناحية، وعامل يمثل المتغير الأساسي المشترك بينهما، من ناحية أخرى. وهذه الارتباطات هي تشعبات البنود على العامل. ويستند معامل أوميغا على افتراضات النموذج المتقارب أو التقاربي congeneric (الذي تمت مناقشته بإيجاز في الفصل الثاني) بدلاً من افتراضات نموذج «تاو» المتماثل أو المترادف في الأساس التي يعتمد عليها معامل ألفا. والاختلاف الرئيسي بين الافتراضات بالنسبة للنموذجين هو أن النموذج التقاربي لا يتطلب أن تكون التباينات للبنود مع عاملها المشترك (الوكيل لتحقيق الدرجة الحقيقية) متساوية أساساً. وكل ما هو مطلوب فعلاً هو أن كل البنود تعكس متغيراً أساسياً مشتركاً واحداً يؤدي إلى عامل مشترك واحد فقط.

ومن الناحية النظرية، فإن استخدام معامل أوميغا يعد أمراً جذاباً للغاية. فهو يعتمد على نفس منطق معامل ألفا، أي تقسيم التباين الملاحظ إلى جزء يعكس الدرجة الحقيقية والباقي يعكس الخطأ. ومن ثم، فإن النسبة الكلية (أي تباين الدرجة الحقيقية زائد تباين الخطأ) التي تعزى إلى الدرجة الحقيقية ستقدر الثبات. ولكنه يفعل ذلك دون قيود ومحددات افتراض لنموذج «تاو» المتماثل أو المترادف في الأساس. ولا يزال افتراض أحادية البعد (أي إن جميع البنود تعكس متغيراً سببياً واحداً مشتركاً) مطلباً، ولكن تم إهمال والتخلي عن القيد الإضافي والمتمثل في أن تكون التباينات بين البنود والدرجة الحقيقية (والتي تقارب احتمالات أو تشعبات البنود على العامل) متساوية بشكل أساسي.

وعلى الرغم من هذه المزايا النظرية، فإنه لم يتم اعتماد معامل أوميغا على نطاق واسع بسبب التعقيد الكبير في طريقة حسابه. على عكس معامل ألفا، والذي يمكن حسابه باستخدام قلم رصاص وورق بمجرد حساب بعض الإحصائيات الأساسية (مثل تباينات البنود والتباين الكلي للمقياس)، أما استخدام معامل أوميغا فإنه ينطوي على زيادة عدد الأرقام بشكل كبير. ولم تكن الإجراءات الروتينية الملائمة لحساب معامل أوميغا متوفرة على نطاق واسع في حزم برامج إحصائية شائعة، ولكن هذا يتغير الآن.

مرة أخرى، يقدم البرنامج الإحصائي R حلاً مجانياً ويمكن الحصول عليها بسهولة للمعضلة الحسابية. على الرغم من أن إدراج وصف مفصل للإجراء التحليلي في البرنامج الإحصائي R هو خارج نطاق هذا الكتاب، فإن "بيتروز" (Peters, 2014) يصف كيفية إعداد البرنامج الإحصائي R بحيث تكون هناك حاجة فقط لكتابة سطرين من التعليمات البرمجية code لحساب كل من معامل ألفا ومعامل أوميغا، كاملة مع حساب فترات للثقة لكل منها. وهذا يسمح للباحث ليس فقط الحصول على معامل أوميغا ولكن أن يقارنه مع معامل ألفا. وبالنسبة لأولئك الذين يفضلون نظرة أكثر تفصيلاً إلى حد ما لعمليات البرنامج الإحصائي R والذين قد يحتاجون إلى بعض التوجيه فيما يتعلق باستيراد ملف بيانات من مصادر مثل إكسل أو برنامج SPSS إلى البرنامج الإحصائي R، فقد وصف دان وآخرون (al.,2014 et Dunn) وصفاً جزئياً عملية آلية مبسطة لحوسبة معامل أوميغا مع فترات الثقة. ومن الجدير بالذكر أن مدخل البوتسترابنج المستخدمة في هذه الحسابات يتضمن إنشاء ما يصل إلى ١٠٠٠ عينة فرعية من مجموعه البيانات الأصلية وأداء الحسابات على كل عينة من هذه العينات الفرعية. وهكذا، فإن العملية برمتها سوف تستغرق وقتاً أطول بكثير حتى تكتمل وتنتهي من الحساب البسيط لمعامل ألفا باستخدام الحزم الإحصائية القياسية. والوقت الذي ينطوي عليه الأمر هو دليل على حقيقة أن هذه الإجراءات أصبحت متاحة على نطاق واسع فقط مع ظهور أجهزة الحاسب الآلي وبيئات البرمجيات المتقدمة مثل البرنامج الإحصائي R.

وفي رأيي، فإن معامل ألفا يحتفظ بالفائدة والمنفعة. وهو كتقدير للحد الأدنى للثبات، فإنه يعد تقديراً متحفظاً للثبات. وعلاوة على ذلك، عندما يتم توخي الحذر في وضع واختيار البنود وأن تكون متوافقة مع افتراض أحادية البعد ونموذج «تاو» المتمثل أو المترادف في الأساس، فإن العديد من المخاطر المحتملة لاستخدام معامل ألفا كمؤشر على الثبات تتضاءل أو تقل. أخيراً، تم استخدام معامل ألفا كرونباخ على نطاق واسع، واستندت معايير الباحثين حول متى تكون أدوات القياس ثابتة بدرجة كافية إلى حد كبير على استخدام معامل ألفا. وإذا كان معامل ألفا متحفظاً بعض الشيء في بعض الحالات، ثم بالنسبة لأي مستوى معين من الثبات، فإن بعض المؤشرات البديلة قد تنتج عنها قيمة رقمية أعلى من قيمة معامل ألفا. قد يكون هذا مربكاً. أي مقياس معين تحت شروط وظروف معينة سيكون له ثبات محدد، ولا تغير طريقة تقدير ذلك الثبات التي تسفر عن قيمة رقمية أكبر من ثبات المقياس. إذا حقق مقياس معين في مناسبة معينة قيمة لمعامل ألفا تساوي ٠,٨٠ وقيمة لمعامل أوميغا بقيمة ٠,٨٢، فإن ثبات المقياس لم يتغير نتيجة للحساب الأخير باستخدام معامل أوميغا.

صحيح أن قيمة معامل أوميغا قد تكون أكثر دقة من الناحية الرقمية، لكن هذا لا يعني أن ثبات المقياس قد تغير، بل إن المقياس المستخدم لقياس هذا الثبات قد تغير. وبعبارة أخرى، فإن القيمة العددية الأعلى البالغة ٠,٨٢، لمعامل أوميغا، في هذه الحالة الافتراضية، تعادل أساساً (أي أنها تمثل نفس درجة الثبات) القيمة ٠,٨٠ لمعامل ألفا. ومن ثم فإن اعتماد وتبني معيار جديد لتمثيل الثبات يتطلب إعادة معايرة معاييرنا الحالية القائمة على أساس معامل ألفا، وهذا لن يكون تحولاً خطياً بسيطاً؛ لا يمكن للمرء أن يحول بين تقدير الحد الأدنى  $g_{lb}$  أو قيمة معامل أوميغا ومعامل ألفا ببساطة بمجرد الإضافة أو الضرب في مقدار ثابت. (في الفصل التالي، أناقش حججاً مماثلة ضد تصحيح الارتباطات الضعيفة عند دراسة الصدق). ولأن معامل أوميغا سيكون أكثر دقة (على سبيل المثال، إذا كان التفكير في معامل الثبات من حيث نسبة التباين في المقياس المشترك مع الدرجة الحقيقية)، فسيكون من المفيد في نهاية المطاف التعود على المقياس الجديد. ومع استمرار العمل على بدائل مختلفة لمعامل ألفا، قد نصل قريباً إلى وقت يتم فيه تبرير مزايا هذه البدائل باعتمادها على نطاق أوسع. ومع ذلك، من المهم في غضون ذلك إدراك أن القيم العددية الأعلى لمعامل أوميغا بالنسبة إلى معامل ألفا لنفس البيانات الناتجة عن نفس المقياس هي إعادة معايرة وليست زيادة في الثبات.

#### الثبات القائم على الارتباطات بين درجات المقياس:

هناك بدائل لمعامل ألفا كمؤشر على الثبات. قد تتضمن هذه الأنواع من حساب الثبات وجود نفس مجموعة الأشخاص الذين يكملون نسختين منفصلتين من مقياس أو نفس النسخة من المقياس في مناسبات متعددة.

#### ثبات النماذج البديلة:

إذا كان هناك شكلان متماثلان بشكل تام للمقياس، يمكن حينئذ حساب الارتباط بينهما طالما أن نفس الأشخاص يكملون الاستجابة على كلا الشكلين المتماثلين. على سبيل المثال، افترض أن الباحث قام أولاً بتطوير مجموعتين متماثلتين من البنود التي تقيس رغبة المرضى في التحكم عند التفاعل مع الأطباء، ثم قام بإدارة وتطبيق مجموعتي البنود لمجموعة واحدة من المرضى، ثم، أخيراً، قام بحساب الارتباط بين درجات مجموعة واحدة من البنود مع درجات مجموعة البنود الأخرى. فإن هذا الارتباط سيكون هو ثبات النماذج البديلة. تذكر أن النماذج المتكافئة أو المتوازية

تتكون من بنود، كلها (إما داخل النماذج أو بينها) تؤدي عملاً جيداً بنفس القدر لقياس المتغير الكامن. وهذا يعني أن كلا الشكليين من المقياس لهما معاملات ألفا، ومتوسطات حسابية، وتباينات متماثلة، ويقيسان نفس الظاهرة. وتتكون النماذج المتكافئة، في جوهرها، من مجموعة واحدة من البنود التي تم تقسيمها بطريقة عشوائية إلى مجموعتين فرعيتين تشكلان النموذجين المتماثلين والبديلين للمقياس. في ظل هذه الظروف، يكون الارتباط بين أحد النماذج والنموذج الآخر مكافئاً أو مساوياً لارتباط أي من النموذجين مع نفسه، حيث إن كل نموذج هو بديل مكافئ للآخر.

#### ثبات التجزئة النصفية:

تتمثل مشكلة النماذج البديلة للثبات في أننا عادةً لا نملك نسختين من المقياس تتوافقان تماماً مع افتراضات الاختبارات المتكافئة أو المتوازية. ومع ذلك، هناك تقديرات أخرى للثبات تطبق نفس النوع من المنطق على مجموعة واحدة من البنود. ونظراً لأن النماذج البديلة تتكون أساساً من مجموعة واحدة من البنود التي تم تقسيمها إلى قسمين، فإنه يترتب على ذلك أنه يمكننا (أ) أخذ مجموعة البنود التي تشكل مقياساً واحداً (أي مقياس لا يحتوي على أي نموذج بديل)، و(ب) تقسيم مجموعة البنود هذه إلى مجموعتين فرعيتين، و(ج) إيجاد الارتباط بين المجموعات الفرعية لتقييم الثبات.

ويُسمى مقياس الثبات من هذا النوع باسم ثبات التجزئة النصفية reliability half-Split. ويُعد ثبات التجزئة النصفية في الواقع تطبيقاً لـ class وليست نوعاً واحداً من الطرق الحسابية نظراً لوجود مجموعة متنوعة من الطرق التي يمكن بها تقسيم المقياس إلى نصفين. وتتمثل إحدى الطرق في مقارنة النصف الأول من البنود بالنصف الثاني. قد يمثل هذا النوع من تقسيم المقياس إلى نصفين هما نصف أول، ونصف أخير إشكالية، لأن عوامل أخرى غير قيمة المتغير الكامن (بمعنى آخر، مصادر الخطأ) قد تؤثر على كل مجموعة فرعية بشكل مختلف. على سبيل المثال، إذا كانت البنود المكونة للمقياس المحدد منتشرة ومتناثرة خلال استبيان مطول، فقد يكون المجيبون أكثر تعباً أو إرهاقاً عند إكمال النصف الثاني من المقياس. وعندئذٍ، سيختلف الإرهاق أو التعب بشكل منتظم بين النصفين، مما يجعلهما أقل تشابهاً على ما يبدو. ومع ذلك، فإن عدم التماثل أو الاختلاف لن يكون سمة من سمات البنود في حد ذاتها بقدر ما هي موضعها في ترتيب البند من المقياس. ومن العوامل الأخرى التي قد تحدث الفرق أو الاختلاف بين البنود في النصف الأول والبنود في النصف

الأخير، تأثير الممارسة (حيث قد تتحسن إجابة المستجيبين على البنود أثناء سيرهم في عملية الإجابة)، أو الإخفاق والفشل في إكمال مجموعه البنود بأكملها، أو ربما حتى شيء عادي مثل التغييرات في جودة طباعة الاستبيان من الأمام إلى الخلف. وكما هو الحال بالنسبة لحالة التعب أو الإجهاد، فإن هذه العوامل ستخفض الارتباط بين النصفين بسبب الترتيب الذي عُرضت به بنود المقياس وليس بسبب نوعيه أو جودة بنود المقياس. ونتيجة لعوامل مثل هذه، فإن قياس قوة العلاقات بين البنود قد يكون معقداً بسبب عدم اتصال الظروف مباشرةً بجودة البند، مما يؤدي إلى تقييم الثبات بصورة خاطئة.

ولتجنب بعض المخاطر المرتبطة بترتيب البند، يمكن للمرء أن يقيم نوعاً آخر من ثبات التجزئة النصفية المعروفة بثبات البنود الفردية والزوجية odd-even reliability. في هذه الحالة، تتم مقارنة المجموعة الفرعية من البنود ذات الترقيم الفردي بالبنود ذات الترقيم الزوجي. وهذا يضمن أن كلاً من المجموعتين الفرعيتين من البنود تتكون من عدد متساو من كل قسم (أي بداية ومنتصف ونهاية) من المقياس الأصلي. على افتراض أن ترتيب البند غير ذي صلة (على عكس الترتيب «السهل إلى الصعب» الشائعة في اختبارات التحصيل، على سبيل المثال)، فإن هذا الأسلوب يتجنب العديد من المشاكل المرتبطة بتقسيمات النصف الأول مقابل النصف الثاني.

ومن الناحية النظرية، هناك العديد من الطرق الأخرى للوصول إلى الثبات بطريقة التجزئة النصفية. هناك بديلان للطرق التي تمت مناقشتها أعلاه لتكوين مجموعات البنود الفرعية هما النصفان المتوازنان balanced halves والنصفان العشوائيان random halves. في الحالة الأولى، يمكن للمرء أن يحدد بعض خصائص أو سمات البند التي يحتمل أن تكون مهمة (مثل، الصياغة في حد ذاتها، أو طول البند، أو ما إذا كان نوع معين من الاستجابة يشير إلى وجود أو غياب السمة المعنية). ثم يتم تشكيل نصفي المقياس بحيث تكون الخصائص ممثلة بالتساوي في كل نصف. وهكذا، يمكن للباحث أن يقسم البنود بحيث يكون لكل مجموعة فرعية نفس عدد البنود التي تمت صياغتها في القسم أو المجموعة الأولى ونفس عدد العناصر القصيرة وما إلى ذلك. غير أنه عند النظر في خصائص البنود المتعددة، قد يكون من المستحيل تحقيق التوازن بين نسبة بند دون أن يكون التوازن بين بند وبند آخر أمراً مستحيلاً. وهذا هو الحال، على سبيل المثال، إذا كانت هناك بنود في القسم أو النصف الأول أكثر من البنود القصيرة. ومن شأن إيجاد توازن لهذه الخاصية



الأخيرة أن يستلزم اختلال التوازن في الخاصية الأولى أو السابقة. أيضاً، قد يكون من الصعب تحديد خصائص البنود التي ينبغي موازنتها.

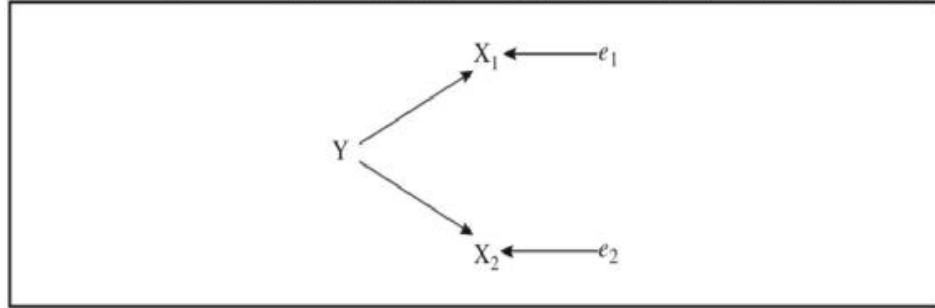
يمكن أن يحصل الباحث على نصفين عشوائيين بمجرد التوزيع العشوائي لكل بند من البنود على إحدى المجموعتين الفرعيتين اللتين سيتم في النهاية حساب الارتباط مع بعضهما البعض لحساب تقدير الثبات. ويتوقف مدى نجاح هذا العمل على عدد البنود وعدد الخصائص موضع الاهتمام، ودرجة الاستقلال بين الخصائص. على أمل أن يسفر عدد قليل من البنود، التي تختلف وفقاً لعدة أبعاد مترابطة، عن مجموعات قابلة للمقارنة من خلال التوزيع العشوائي، وهذا أمر غير واقعي. ومن ناحية أخرى، فإن التوزيع العشوائي لمجموعة مكونة من 50 بنداً تتفاوت وتختلف فيما يتعلق بسمتين أو ثلاث سمات غير مترابطة إلى فئتين قد يؤدي إلى مجموعات فرعية قابلة للمقارنة فيما بينها بشكل معقول.

وتعتمد أفضل طريقة يتم بها الحصول على نصفين مجزأين على الموقف أو الحالة الخاصة بها. وأهم ما في الأمر هو أن يفكر الباحث في طريقة تقسيم البنود التي يمكن أن تؤدي إلى مجموعات فرعية غير متكافئة والخطوات التي يمكن اتخاذها لتجنب ذلك. إن التبرير المنطقي وراء كل من ثبات التجزئة النصفية وثبات النماذج البديلة هو امتداد طبيعي لنموذج الاختبارات المتكافئة (المتوازية).

وعلى الرغم من أننا عندما ناقشنا هذا النموذج في البداية، اعتبرنا أن كل بند هو اختبار، يمكن للمرء أن ينظر أيضاً إلى مقياس (أو نصفين من مقياس) يتوافق مع النموذج كاختبار. ولذلك، يمكننا تطبيق المنطق الذي استخدمناه في حالة البنود المتعددة على حالة اثنين من النماذج البديلة أو نصفين من المقياس. فكر في اختبارين (نصفي المقياس أو النماذج البديلة) تحت افتراضات الاختبارات المتكافئة، كما هو موضح في الشكل رقم (٣-٣).

والسبيل الوحيد الذي يربط بين المكونين الاثنين للمقياس هو المسارات السببية من المتغير الكامن إلى كل نصف من نصفي المقياس. وهكذا، فإن ناتج قيم هذه المسارات يساوي الارتباط بين الاختبارات. وإذا كانت قيم المسار لا بد أن تكون متساوية (وهي تكون كذلك تحت افتراضات هذا النموذج)، فإن الارتباط بين الاختبارات يساوي مربع قيمة المسار من المتغير الكامن إلى أي من الاختبارين. مربع ذلك المسار (على افتراض أنه معامل مسار معياري) هو أيضاً نسبة التباين في

أي اختبار يتأثر بالمتغير الكامن. وهذا بدوره هو تعريف الثبات. وبالتالي، فإن الارتباط بين الاختبارين يساوي ثبات كل منهما.



شكل (٣-٣): مخطط مسار يوضح علاقات نصف المقياس منفصلين (X1) و (X2) بالمتغير الكامن المشترك بينهما

وفي حين أن الاختبارات المشار إليها في الفقرة السابقة هي عبارة عن نسختين كاملتين من المقياس في حالة النماذج البديلة، إلا أنهما عبارة عن نصفين للمقاييس في حالة التجزئة النصفية. وبالتالي، فإن الارتباط بين نصف المقياس يؤدي إلى تقدير الثبات لكل نصف من المجموعة الكاملة من البنود، وهو ما يقلل من تقدير ثبات مجموعة البنود بأكملها. ويمكن حساب تقدير ثبات المقياس بأكمله، استناداً إلى ثبات جزء من المقياس، باستخدام معادلة سبيرمان براون-Spearman-Brown formula، التي تمت مناقشتها سابقاً في هذا الفصل. تذكر أنه وفقاً لهذه المعادلة، فإن معامل الثبات ألفا يساوي:

$$\alpha = \frac{k\bar{r}}{1 + (k - 1)\bar{r}}$$

حيث  $k$  هي عدد البنود المعنية وهو متوسط ارتباط بين بند ما مع أي بند آخر (أي، متوسط الارتباط بين البنود). إذا كنت قد حددت مدى ثبات مجموعة فرعية من البنود (على سبيل المثال، من خلال طريقة التجزئة النصفية) وتعرف عدد البنود التي يستند إليها الثبات (على سبيل المثال، نصف العدد في المقياس بأكمله)، فيمكنك استخدام المعادلة لحساب متوسط ارتباط بين بند ما مع أي بند آخر. بعد ذلك، يمكنك بمساعدة قيمة هذه وعدد العناصر في المقياس بأكملها لتعويض في المعادلة. وستكون النتيجة تقديراً لثبات المقياس بالكامل، استناداً إلى قيمة الثبات المحسوبة لنصفي

المقياس المجزأ. وهذا يعمل على تبسيط الأمور إذا قمت بإجراء بعض العمليات الجبرية (الحسابية) على معادلة سبيرمان - براون لوضعها في الشكل التالي:

$$r = \frac{r_{yy}}{[k - (k - 1) r_{yy}]}$$

حيث  $r_{yy}$  هو ثبات البند المعني. على سبيل المثال، إذا علمت أن ثبات التجزئة النصفية لنصفي مقياس مكونين من تسعة بنود تساوي ٠,٩٠، يمكنك حساب  $r_{yy}$  على النحو التالي:

$$\bar{r} = \frac{.9}{[9 - (8)(.9)]} = .5$$

ويمكنك بعد ذلك إعادة حساب الثبات للمقياس الكامل المكون من 18 بنوداً باستخدام تساوي (٠,٥)، وعدد البنود  $k$  يساوي (١٨) في معادلة سبيرمان - براون. وعلى ذلك، فإن تقدير الثبات للمقياس الكلي يكون:

$$\frac{18 \times .5}{1 + (17 \times .5)}$$

الذي يساوي 9 / 5.9، أو 947.0. لاحظ أن زيادة عدد البنود قد زاد من الثبات. إن إلقاء نظرة سريعة على معادلة سبيرمان - براون من شأنه أن يجعل من الواضح أنه، إذا كان كل شيء متساوياً، فإن المقياس الأطول سيكون دائماً أكثر ثباتاً من المقياس الأقصر. قد لا يكون هذا واضحاً بشكل حدسي أو بديهي. إذا كان لكل بند بعض الإشارات وبعض الضوضاء، ألا تؤدي إضافة البنود إلى زيادة الضوضاء وكذلك الإشارة؟ الجواب هو، نعم، سوف تزيد البنود ذلك؛ ومع ذلك، يبين لورد ونوفيك (Lord and Novick, 2008) أنه بينما يزداد الخطأ خطئاً، تزداد الإشارة signal أضعافاً مضاعفة عند إضافة البنود (ص. 86) – أي إن تباين الدرجة الحقيقية يتراكم بشكل أسرع من تباين درجة الخطأ مع إطالة الاختبار. وهذا مماثل للطريقة التي يصبح بها أي تقدير لمعلمة المجتمع (على سبيل المثال، متوسط وزن المراهقين في الولايات المتحدة) تدريجياً أكثر دقة (أي أقرب إلى الدرجة الحقيقية للمجتمع) مع زيادة عدد الأفراد الذين تم أخذهم في العينة.

## اتفاق المُقيِّمين inter-rater agreement:

تضمنت أساليب تقييم الثبات التي ناقشناها حتى الآن على البنود كمؤشرات. ومن المؤكد أن هذا مناسب لكتاب عن وضع وتطوير المقياس؛ غير أنه، في العديد من السياقات البحثية، يعمل المقيمون أو المحكمون كمؤشرات. على الرغم من عدم وجود صلة وثيقة لذلك بتطوير المقياس، إلا أن مناقشة موجزة لاتفاق المقيمين (انظر دي فيليس، 2005 DeVellis، للحصول على نظرة عامة حول هذا الموضوع) تكون مفيدة لأنها يمكن أن تؤكد التشابه المفاهيمي الذي تشترك فيه مع تقييم الثبات القائم على البند. ففي كلتا الحالتين، يفرض المنطق الأساسي أن المؤشرات التي تتقاسم قضية مشتركة يجب أن ترتبط مع بعضها البعض. في حالة المقيمين أو المحكمين، إذا كانت الدرجات الناتجة عن ملاحظاتهم تعكس خصائص المثير أو المنبه الذي يتم ملاحظته بدلاً من خصائص المحكمين، فينبغي عندئذ أن تتفق تلك الدرجات. ويمكن أن تختلف طبيعة هذا الاتفاق تبعاً لأهداف الباحث والمنهج المتبع في تقدير اتفاق المقيمين.

وفي بعض الحالات، سيحتاج الباحث إلى اتفاق صارم من أجل النظر في التقديرات التي يمكن التعويل عليها والوثوق بها. وكمثال على ذلك، فإن المدى الذي يمكن فيه لاثنتين من حكام البيسبول (كرة القدم الأمريكية) أن يحددا بشكل ثابت متى يكون ضرب الكرة صحيحاً أو خطأ يتوقف على مدى اتفاقهما على سلسلة من الأحداث أو المناسبات. يجب أن يكون أي ضرب للكرة إما صحيحاً أو خطأ وفقاً لقواعد البيسبول. ومن ثم، فإن الاتفاق المطلق يكون كافياً كوسيلة لقياس مدى ثبات تقديرات الحكام. وفي ظروف أخرى، قد يقوم المقيمون بتقييم ظاهره تختلف على طول سلسلة متصلة. على سبيل المثال، من المحتمل أن يستخدم الحكام في سوق ولاية ما يقومون بتقييم نوعية أو جودة الفطائر المخبوزة محلياً مقياساً له العديد من الاستجابات بدلاً من التقسيم الثنائي البسيط (جيد أو سيئ). وفي مثل هذه الحالة، قد لا يكون من الضروري التوصل إلى اتفاق دقيق حول الدرجة لتحديد أن تقديرات الحكام المتعدين تعكس في الواقع سمة الفطائر وليست مجرد سمات الحكام أنفسهم. وعلى الرغم من أن أحد الحكام قد يكون أكثر تساهلاً من قاض آخر غيره وبالتالي يعطي تقييمات أعلى باستمرار، إلا أن الدرجات بين الحكام ينبغي أن تتماثل تقريباً إذا كان يجب اعتبارها تتمتع بالثبات. والواقع، قد يكون هذا الأساس الأفضل للحكم على الثبات في هذه الظروف بدلاً من الإصرار على الاتفاق التام. والتقييمات التي تم تعيينها حتى يكون ترتيب الفطائر المحرز أو المسجل من الأفضل إلى الأسوأ متشابهاً بين الحكام توحى بأن سمات الفطائر وليست

سمات الحكم هي التي تحدد الدرجات التي تم تعيينها. ومع ذلك، إذا كانت معايير تسجيل الحكم تمت معاييرها بشكل مختلف إلى حد ما (مثلاً، إذا اعتبر أحد الحكم أن الدرجة خمسة تمثل متوسط جودة الفطيرة، بينما أعطى آخر للفطيرة متوسطة الجودة ست درجات)، فقد لا يتفق الحكم أبدًا في الواقع. ومن ثم، قد لا يكون الاتفاق التام معيارًا مناسبًا لتقييم الثبات فيما بين الحكم في مثل هذه الحالة.

وفي ورقه مؤثره للغاية، وصف باتريك ا. شروت وجوزيف ل. فليس (Patrick E. Joseph and ShROUT L. Fleiss, 1979) الطرق المختلفة لتقييم اتفاق المقيمين في ظل ظروف مختلفة. وأكدوا على أن السمات المهمة لنوايا الباحث تحدد أفضل طريقة لتقييم الثبات فيما بين المقيمين. ولأغراض المناقشة الحالية، فإن أهم سمة ناقشوها هي ما إذا كان الاتفاق الدقيق أو مجرد التطابق بين المحكمين يمثل أفضل دليل على اتساق المقيمين. وفي مثال الحكم، من الواضح، أن الاتفاق المطلق سيكون ضرورياً لاستنتاج أي من الحكم يتمتع بالثبات. إن نسخة معامل الارتباط داخل الطبقة ICC المناسبة في هذه الحالة تعادل أو تساوي معامل كوهن كبا (Cohen's kappa (s (كوهين؛ Cohen, 1960). ويحدد هذا المدخل إلى أي مدي يتجاوز تكرار الاتفاقات الدقيقة بين المحكمين ما يمكن توقعه عن طريق الصدفة. وعلى النقيض من الحكم، يمكن لقضاه الحكم على جودة الفطيرة أن يبرهنوا على أنهم كانوا يقومون بإصدار أحكام تتمتع بالثبات دون الاتفاق الفعلي على ذلك بدقة. نموذج معامل الارتباط داخل الطبقة ICC الذي يعبر عن الاتفاق في ظل هذه الظروف يعادل معامل ألفا كرونباخ.

وعلى ذلك، يمكن استخدام معامل ألفا كرونباخ عندما يهتم المرء بالمدى الذي يسفر فيه أي من البنود أو المقيمين عن درجات تكون متوافقة أو متطابقة. وفي كلتا الحالتين، ينشأ التوافق أو التطابق (أي الارتباطات بين البنود وبعضها البعض أو بين المقيمين) بسبب التأثير المشترك للمتغير الأساسي على المؤشرات، سواء كانت هذه الأخيرة عبارة عن بنود أو مقيمين. وتشكل نسبة تباين الدرجة الملاحظة التي يمكن إرجاعها إلى الدرجة الحقيقية للمتغير الكامن ثبات المؤشرات في كلتا الحالتين. وتؤكد هذه المقارنة بين أنواع مختلفة من المؤشرات (المقيمين مقابل البنود) تفسير الفائدة العامة لمعامل ألفا كرونباخ على الثبات.

الاستقرار الزمني:

هناك طريقة أخرى لحساب الثبات تنطوي على الاستقرار الزمني للقياس stability temporal، أو مدى بقاء الدرجات أو النتائج ثابتة (مستقرة) من مناسبة أو من حالة إلى أخرى. ويُعد ثبات إعادة الاختبار Test-retest reliability هي الطريقة المستخدمة عادةً لتقييم ذلك النوع من الثبات. وفي بعض النواحي، يكون ثبات إعادة الاختبار مماثلاً ومشابهاً لثبات التجزئة النصفية وثبات النماذج البديلة، والتي تمت مناقشتها سابقاً. وبالعودة إلى مثال سابق، افترض أنه بدلاً من تطوير مجموعتين من البنود لقياس رغبة المرضى للتحكم أو السيطرة عند التفاعل مع الأطباء، وضع باحثنا الافتراضي مجموعة واحدة من البنود. ويمكن تطبيق هذه البنود على مجموعة واحدة من المرضى في مناسبتين منفصلتين، ويمكن حساب ارتباط الدرجات التي حصلوا عليها في المناسبة الأولى بتلك الدرجات التي حصلوا عليها في المناسبة الثانية. والأساس المنطقي الذي تستند إليه القرارات المتعلقة بالثبات من هذا النوع هو أنه إذا كان المقياس يعكس حقاً (في الواقع) مغزى لبناء مفاهيمي محدد، فإنه يجب عليه تقييم ذلك المفهوم البنائي بطريقة مماثلة أو متساوية في مناسبات أو حالات منفصلة. وبعبارة أخرى، فإن الدرجة الحقيقية للمتغير الكامن ينبغي أن تمارس تأثيراً مشابهاً أو مماثلاً على الدرجات الملاحظة في مناسبتين (أو أكثر)، في حين ينبغي ألا يظل مكون الخطأ ثابتاً عبر تطبيقات المقياس. وبالتالي، ينبغي أن يمثل الارتباط بين الدرجات التي تم الحصول عليها عبر تطبيقين للمقياس على نفس الأفراد المدى الذي يحدد به المتغير الكامن الدرجات الملاحظة. وهذا يعادل تعريف الثبات كنسبة التباين التي تعزى إلى الدرجة الحقيقية للمتغير الكامن.

إن مشكلة هذا المنطق تكمن في أن ما يحدث للدرجات بمرور الزمن قد يتعلق بتعرض إجراءات القياس للخطأ أو لا يكون له صلة بذلك. ويشير «نونالي» (Nunnally, 1978) إلى أن خصائص أو سمات البنود قد تجعلها تنتج ردوداً مستقرة مؤقتاً حتى عندما يتغير المفهوم البنائي محل الاهتمام. على سبيل المثال، إذا تأثر مقياس القلق المزعوم (المفترض) بالرغبة الاجتماعية وكذلك القلق، فقد تظل الدرجات ثابتة على الرغم من الاختلافات أو التباينات في القلق. لن يكون الاستقرار في الدرجات، الذي ينعكس في الارتباط المرتفع بين مناسبات تطبيق المقياس، نتيجة الثبات في الظاهرة موضع الاهتمام. وبدلاً من ذلك، قد لا تتغير هذه الظاهرة بينما تتغير الدرجات في المقياس، أي إن المقياس قد يكون غير ثابت. أو قد تُعزى التغيرات في الدرجات إلى عدم الثبات عندما تتغير الظاهرة نفسها في الواقع، وقام المقياس بتعقب هذا التغير بشكل دقيق.

والمشكلة هي أن التغيير أو عدم التغيير يمكن أن يُعزى إلى مجموعة متنوعة من الأشياء إلى جانب ثبات (أو عدم ثبات) إجراءات القياس. حدد "كليي وماكجراث" (Kelly and McGrath, 1988) أربعة عوامل مربكة عندما يفحص المرء مجموعتين من الدرجات لنفس المقياس، بينهما فاصل زمني. وهذه العوامل هي (1) التغيير الحقيقي في المفهوم البنائي محل الاهتمام (على سبيل المثال، صافي الزيادة في متوسط مستوى القلق بين عينة من الأفراد)، (2) التذبذبات أو التقلبات المنتظمة في الظاهرة (على سبيل المثال، التباينات في القلق حول متوسط ثابت كدالة في الزمن)، (3) التغييرات التي تعزى إلى الاختلافات في الموضوعات أو طرق القياس بدلاً من الظاهرة محل الاهتمام (مثل تأثيرات التعب والإجهاد الذي يتسبب في حدوث أخطاء في قراءة البنود)، و(4) عدم الاستقرار الزمني بسبب عدم الثبات المتأصل في إجراءات القياس. والعامل الرابع فقط هو عدم الثبات.

ومؤخراً، ناقش "يو" (Yu, 2005) كيف أن الأخطاء في درجات ثبات إعادة الاختبار لا تنشأ فقط من أوجه القصور في الأداة نفسها ولكن أيضاً تنتج من الأفراد الذين يتم تطبيق المقياس عليهم والإجراءات الخاصة بتطبيق الأداة. وتشمل الأخطاء الناشئة عن المشاركين الذين يجري تقييمهم آثار الانتقال التي يؤثر فيها الأداء في الاختبار الأول على الأداء في الاختبار اللاحق. على سبيل المثال، قد يتذكر المجيب إجاباته السابقة وقد يرغب في أن يظهر بشكل متسق أو ثابت. وبالتالي، فإن الإجابات على التطبيق الثاني للاختبار لا تحددها بشكل مباشر حالة المتغير محل الاهتمام، بل يحددها الحافز على الظهور بشكل متسق مع مرور الزمن. يشير «يو» إلى مصدر خطأ آخر يرجع إلى المجيب – وهو تأثير التشاور أو المناقشة. وكمثال، قد يأخذ طلاب المدارس الثانوية في البداية اختبار القدرة الدراسية أساساً كوسيلة لاكتساب الخبرة في عملية الامتحان. ويمكنهم بعد ذلك التشاور والمناقشة بشأن هذه التجربة واستخدامها كأساس للتحضير لاختبار لاحق. وكمثال آخر لتأثير التشاور أو التداول، يمكن للمشاركين الذين يجدون أسئلة هجومية في التطبيق المبدئي لأداة ما أن يعتمدوا حذف أو تشويه إجاباتهم عندما يتم إعادة النظر في البنود وقراءتها. إن أخطاء التطبيق التي ذكرها «يو» تتضمن اختلافات في الإجراءات بين الاختبار وإعادة الاختبار (على سبيل المثال، السماح للمستجيبين بالعودة والتحقق من الإجابات السابقة)، والتعليمات غير الجيدة، والذاتية في تسجيل الإجابات (على سبيل المثال، المعايير الغامضة الخاصة بتسجيل

(الإجابات)، والفترة الزمنية غير المناسبة بين تطبيقات الاختبار. ويمكن لأي من هذه الأنواع من الأخطاء أن تقلل من درجات الثبات دون أن تكون الأداة نفسها غير ثابتة بطبيعتها.

وهذا لا يعني أن إثبات الاستقرار الزمني أمر غير مهم. في أي عدد من سياقات البحث، قد يكون من الضروري افتراض (أو إثبات) أن القياسات المنفصلة أي التي توجد بينها فترة زمنية ترتبط ارتباطاً وثيقاً (عالياً). ومع ذلك، فإن الاستقرار الذي نسعى إليه في هذه الحالات هو استقرار كل من القياس والظاهرة. وتُخبرنا ارتباطات إعادة الاختبار عن القياس فقط عندما نكون على ثقة كبيرة من أن الظاهرة ظلت مستقرة ولم تتغير. وهذه الثقة لا يوجد ما يبررها في كثير من الأحيان. وهكذا، فإن أفضل طريقة لثبات إعادة الاختبار، على الرغم من أهميتها، هي الكشف عن شيء ما حول طبيعة الظاهرة وقياسها، وليست الأخيرة (أي القياس) وحدها. وبالإشارة إلى الثبات أو عدم التغير في الدرجات بمرور الزمن فإن الاستقرار الزمني هو الأفضل لأنه لا يشير، كما هو الحال بالنسبة لثبات إعادة الاختبار، إلى أن خطأ القياس هو مصدر أي عدم استقرار نلاحظه.

إن حساب ثبات إعادة الاختبار عادةً ما ينطوي على شكل ما من أشكال معامل الارتباط. والاختباران اللذان يفصل بينهما فترة زمنية يشبهان ويمثلان مجموعتين من البنود يتم تطبيقهما في نفس الوقت. وفي الحالة الأخيرة، وكما ذكر آنفاً، فإن الارتباط بين البنود يعكس حاصل ضرب المسارين الممتدين من المتغير الكامن إلى البنود. ووفقاً لافتراضات الاختبارات المتكافئة التامة، تكون لهذه المسارات قيم متطابقة ولذلك فإن حاصل ضربها تساوي القيمة التربيعية لكل منها. كما يمثل كل مسار الارتباط بين المتغير الكامن عند طرفه الأصلي والمؤشر عند طرفه النهائي. ومن ثم، فإن مربع هذه القيمة هو الارتباط التربيعي، أو نسبة التباين المشترك بين المتغير الكامن والمؤشر. لذلك، فإننا من خلال الربط بين البندين، نحصل على القيمة التي هي من الناحية المفاهيمية تعادل أو تساوي (في إطار افتراضات الاختبار المتكافئ التام) نسبة التباين المشترك بين البند والمتغير الكامن. وفي حالة ثبات إعادة الاختبار، يكون المنطق مطابقاً لذلك ولكن المؤشرين هما الآن الأداة التي تم تطبيقها في الوقت الأول ثم تم تطبيقها مره أخرى في الوقت الثاني. وينبغي أن يكون ارتباطهما هو تقديراً للارتباط التربيعي (وبالتالي، نسبة التباين المشترك) بين الدرجة التي تم الحصول عليها (من أي من التطبيقين للأداة) والدرجة الحقيقية للمتغير الكامن. وعندما لا يتم الوفاء بافتراضات الاختبارات المتكافئة التامة، تصبح القيمة الناتجة حداً أدنى للثبات الحقيقي

lower bound.



ويشير هذا التحليل إلى أنه يمكن تقدير ثبات مجموعه من البنود يتم تطبيقها في مناسبتين أو فترتين زمنيتين من خلال الارتباط بين الدرجات المتحصل عليها من هذين التطبيقين. وهذا، من الناحية العملية، هو طريقة حساب ثبات إعادة الاختبار عادة. في الظروف الشائعة التي تكون فيها الدرجات الموجودة في أداة الاختبار من نوع القياس المستمر أو المتصل أساساً، يتم عادةً استخدام ارتباط بيرسون لحاصل ضرب العزوم Pearson product-moment correlation. وينبغي أن يعمل ارتباط بيرسون بشكل جيد في معظم الحالات الخاصة بأدوات القياس التي تم تطويرها باستخدام المنهج الموضح في هذا الكتاب. وإذا كان الباحث قلقاً من أن الدرجات المعنية لا تتطابق مع المقياس من نوع القياس الفتري أو المسافي (على سبيل المثال، القياس الترتيبي)، فإنه يمكن الاستعاضة بشكل بديل عن معامل الارتباط «بيرسون». ومع ذلك، لاحظ أنه حتى لو لم يكن كل بند من البنود على حدة من نوع القياس الفتري أو المسافي التام، فإن الدرجة القائمة على جمع هذه البنود معاً سوف تكون متقاربة إلى حد بعيد للبيانات المقاسة على مقياس فتري أو مسافي.

#### ثبات درجات التغيير:

هناك جانب آخر يستحق منا الاهتمام عند النظر في الثبات عبر الزمن. ونظراً لأن المقياس يظهر أنه يتمتع بالثبات، فإن هذا لا يتبعه بالضرورة أن درجات التغيير (المعروفة أيضاً باسم درجات الفرق أو الاختلاف)، والتي يتم حسابها على أنها الفرق بين درجات تطبيقين لأداة القياس يفصل بينهما فترة زمنية، ستتمتع بالثبات بنفس القدر. في الواقع، لاحظ العديد من المؤلفين (على سبيل المثال لين، سليندي (1977)؛ أوفرال، وودوارد (1975)؛ Linn & Slinde, 1977; Overall & Woodward, 1975) أن درجات التغيير تميل إلى أن تكون أقل ثبات من درجات مقياس تم تطبيقه مرة واحدة وتم الحصول على الدرجات من نفس أداة القياس.

وقبل النظر في هذا لموضوع ودراسته بشكل أكبر، دعونا نفكر فيما تمثله ثبات درجة التغيير. يفترض مفهوم الثبات بالنسبة لدرجة التغيير أو الفرق أن: (أ) هناك تباين في التغيير و(ب) أن جزءاً من هذا التباين يعزى إلى فروق حقيقية في التغيير بينما يعزى الجزء الباقي إلى عوامل أخرى (أي الخطأ). ومن المفترض، أن الأفراد في موقف معين تكون لهم «درجات حقيقية» مختلفة تمثل تغيراتهم الفردية بمرور الزمن. أي إن القيمة الحقيقية المقابلة للتغيير في الزمن ستختلف من فرد إلى فرد آخر. إلى الحد الذي يتأثر فيه التباين بين الأفراد في درجات

التغيير بالاختلافات الفعلية في الدرجات الحقيقية بدلاً من الخطأ، فإن قياس التغيير هذا يكون ثابتاً. لنأخذ مثلاً مبالغاً فيه (متطرف)، إذا تغير كل شخص بنفس المقدار تماماً، فلن توجد اختلافات فردية في الميل إلى التغيير. أي إنه لا يوجد اختلاف بين الأفراد في التغيير الحقيقي، وبالتالي، لا يوجد أي ارتباط بين التغيير الحقيقي والتغيير الملحوظ (كما هو الحال دائماً عندما يكون أحدهم ثابتاً). وبالتالي، يجب أن تكون قيمة ثبات درجة التغيير (أي الارتباط التربيعي بين الدرجات الحقيقية والدرجات الملحوظة) تساوي الصفر أيضاً. وبديهياً، إذا تغير الجميع في انسجام تام، فمن المحتمل أن نفترض أن التغيير يمثل نوعاً من التأثير الظرفي بدلاً من السمات الشخصية البحتة. إذا كان الشيء الذي يهمنا قياسه هو سمة شخصية، فإن درجة التغيير التي تعكس في المقام الأول ظاهرة موقفية وليست الظاهرة الشخصية هدفنا.

كيف تبدو معادلة ثبات درجات التغيير؟ وكما هو الحال مع أي معامل ثبات، فإن نسبة التباين الكلي يمكن أن تعزى إلى التباين الحقيقي. وهكذا:

$$r_{\Delta} = \frac{\sigma_{\Delta T}^2}{\sigma_{\Delta_0}^2} = \frac{\text{تباين الدرجة الحقيقية}}{\text{تباين الدرجة الملحوظة}}$$

حيث  $(r_{\Delta})$  هي ثبات درجة التغيير، و  $(\sigma_{\Delta T}^2)$  هي تباين درجة التغيير الحقيقية، و  $(\sigma_{\Delta_0}^2)$  هي تباين درجة التغيير الملحوظ (أي مجموع تباينات الدرجة الحقيقية والخطأ). ويمكننا التعامل مع تباينات درجات التغيير كأن لدينا تباينات أخرى من التوليفات، أي تساوي مجموع كل تباينات المكونات، بالإضافة إلى ضعف تغايراتها. وهذا مماثل لتوسيع النطاق أو المقدار الجبري  $(X + Y)^2$  إلى  $X^2 + Y^2 + 2XY$ .

ومع ذلك، نظراً لأننا نتعامل مع درجات الفروق على الشكل  $D = X_2 - Y_1$  فإننا نقوم بعملية طرح وليست عملية إضافة أو جمع، لذلك فإن المقدار الجبري هو في الواقع مماثل للتوسيع  $(X + Y)^2$  into  $X^2 + Y^2 + 2XY$  ولأن  $(X_2 - X_1)^2 = X_1^2 + X_2^2 - 2(X_1 * X_2)$ ، عندما نوسع التباين لدرجة الفرق أو التغيير، فإننا نصل إلى:

$$r_{\Delta} = \frac{\sigma_{T_{X1}}^2 + \sigma_{T_{X2}}^2 - 2\sigma_{X1,X2}}{\sigma_{X1}^2 + \sigma_{X2}^2 - 2\sigma_{X1,X2}}$$

حيث يكون المقدران الأوليان في البسط هما تباينات الدرجات الحقيقية لأداة القياس الأولية وأداة القياس اللاحقة، على التوالي، للمقياس موضع الاهتمام؛ والمقدار الأخير في البسط هو تبايناتهم. تتوافق مقادير المقام مع نظيراتها في البسط، لكن التباين يكون للدرجات الملاحظة وليس للدرجات الحقيقية.

يمكننا تقدير تباينات الدرجات الحقيقية من خلال حاصل ضرب ثبات المقاييس في تباينات درجاتها الملاحظة. وهذا، في جوهره، يعطينا نسبة تباين الدرجات الملاحظة التي نعتبرها ناتجة أو ناشئة من الدرجات الحقيقية بدلاً من الخطأ. باستخدام هذا الإجراء لتقدير التباين الدرجات الحقيقية، يكون المقدار السابق مساوياً لـ:

$$r_{\Delta} = \frac{[(\sigma_{X1}^2) \times (r_{X1,X1})] + [(\sigma_{X2}^2) \times (r_{X2,X2})] - [2\sigma_{X1,X2}]}{(\sigma_{X1}^2) + (\sigma_{X2}^2) - (2\sigma_{X1,X2})},$$

حيث  $(r_{X1,X1})$  و  $(r_{X2,X2})$  هي ثبات القياسات الأولى والثانية على التوالي. لاحظ أنه في كلتا المعادلتين السابقتين، يكون التباين (التباين المشترك) متطابقاً في كل من البسط والمقام. ويرجع ذلك إلى أنه إذا افترضنا أن الأخطاء غير مرتبطة عبر الزمن، فإن كل التباين عبر الزمن هو تباين يتميز بالثبات. وهذا يعني أن التباين في الدرجات الملاحظة لا يمكن أن ينشأ إلا عن التباين للدرجات الحقيقية إذا استمر افتراض عدم ارتباط الأخطاء قائماً. وبالتالي، فإن التباين الملاحظ والتباين الحقيقي يكونان نفس الشيء (متساويان).

باستخدام المعادلة السابقة، يمكننا عمل جدول لماهية ثبات درجة التغيير التي ستكون لقيم مختلفة لتباينات الدرجات والتباينات. لنفترض أن لكلا الدرجات الناتجة عن تطبيق الأداة في الوقت الأول والوقت الثاني تباين الوحدة وأن للمقياس موضع الاهتمام نفس الثبات في كل من الوقت الأول والوقت الثاني. يمكننا الآن أن نرى كيف يؤثر التباين الخاص بهما على ثبات درجة التغيير.

$\sigma_{X1,X2}$	ثبات درجة التغير (rΔ) عندما تكون درجة الثبات		
	$r_{X1,X1} = r_{X2,X2} = .50$	$r_{X1,X1} = r_{X2,X2} = .70$	$r_{X1,X1} = r_{X2,X2} = .90$
.00	.50	.70	.90
.10	.44	.67	.89
.20	.38	.63	.88
.30	.29	.57	.86
.40	.17	.50	.83
.50	.00	.40	.80
.60	—	.25	.75
.70	—	.00	.67
.80	—	—	.50
.90	—	—	.00

ما نراه هو أن الثبات الأعلى في وقت واحد يرتبط بزيادة ثبات درجة التغير (أي إن الزيادة في قيم ثبات درجة التغير من اليسار إلى اليمين عبر أي صف من الجدول)، وهو أمر غير مفاجئ أو مثير للدهشة. ما قد يكون أكثر إثارة للدهشة في البداية هو أن التباين الأعلى للدرجات عبر الزمن يؤدي إلى انخفاض ثبات درجة التغير (أي انخفاض قيم ثبات درجة التغير من أعلى إلى أسفل على طول أي عمود من الجدول).

يشير روجوسا (1988) Rogosa، الذي كان من بين أقوى المؤيدين لدرجات التغير، إلى أن الجداول التي تعطي قيم الثبات فقط للدرجات التي ترتبط ارتباطاً كبيراً عبر الزمن (مثل تلك الموجودة في الصفوف السفلية من الجدول السابق) قد ترسم صورة قاتمة بشكل لا مبرر لها لثبات درجات التغير. ويجادل «روجوسا» بأن التباين العالي في الدرجات عبر الزمن يقيد أو يحد من نسبة التباين الملاحظ في الوقت الثاني والذي يمكن أن يعزى إلى الاختلافات الفردية في التغير. والسبب في التباين المرتفع مع مرور الزمن هو أن كل شخص يتغير بشكل أساسي في انسجام تام بدلاً من أن يكون هناك تباين كبير بين شخص وشخص آخر في تغيير الدرجات بمرور الزمن. إذا كان هناك تباين أو اختلاف بسيط في المعدل الحقيقي للتغير، فلا يمكن أن يكون هناك الكثير من التباين بين درجة التغير الحقيقية ودرجة التغير الملاحظة (تذكر أنه لكي يكون هناك تباين كبير يجب أن يكون هناك تباين كبير لكلا المتغيرين). ويذهب «روجوسا» (1988) إلى القول بأنه، عندما يكون المدى الحقيقي للتباين في معدل التغير كبيراً، فإن ثبات درجات التغير بالنسبة لثبات

الدرجات في نقطة زمنية واحدة تكون مواتية بدرجة كبيرة. توضح الصفوف العليا من جدولنا هذه النقطة. ومع ذلك، فإن الجداول المنشورة في كثير من الأحيان لم تقدم إلا الظروف القصوى المبينة في الصفوف السفلية من الجدول السابق.

ماذا يعني كل هذا على المستوى العملي أو التطبيقي؟ من المتوقع أن تؤدي الظروف التي تحدث فيها الأحداث على مر الزمن إلى تأثير مماثل من حيث الحجم على جميع المشاركين، مما سيؤدي إلى ضعف ثبات درجات التغيير لأنه لن يكون هناك تباين يذكر بين الناس. وهذا يعني أنه إذا كانت حصة الأسد (الحصة الأكبر) من التباين ناتجة عن عوامل تُعزى إلى مرور الزمن (مثل تلقي تدخل يؤثر على جميع المشاركين بالتساوي)، فإن الحصة التي تُعزى إلى الاختلافات الفردية يجب أن تكون صغيرة نسبياً في المقابل. إذا زادت درجات كل الأشخاص من الوقت الأول إلى الوقت الثاني بمقدار خمس نقاط، على سبيل المثال، فيمكنك أن تتنبأ أو تتوقع بدقة درجة الوقت الثاني من خلال الدرجة السابقة. وسيكون ارتباط الدرجات في الوقتين هو واحد صحيح، ودرجة التغيير الخام ستكون ثابتة (أي خمس نقاط لكل الأشخاص)، وبالتالي فإن ثبات درجة التغيير ستكون صفراً. وعلى العكس من ذلك، فإن الحالات التي يرجح أن تختلف فيها التغييرات بمرور الزمن اختلافاً كبيراً من فرد إلى آخر تخلق على الأقل فرصة لكي تكون درجات التغيير أكثر ثباتاً. وإذا نظرنا إليها من منظور مختلف قليلاً، إذا كان مقدار التباين قد ساهم في درجات الوقت الثاني من خلال درجات الوقت الأول يبلغ 100٪ تقريباً، فإن مقدار التباين في درجة الوقت الثاني لا يمكن أن تنسب إلى متغيرات أخرى غير درجة الوقت الأول. يتم تحديد درجة الوقت الثاني أساساً من خلال الدرجات التي تم تسجيلها، بغض النظر عما قد حدث بين الوقت الأول والوقت الثاني. ومن ثم، فإن مؤشر التغيير المفترض في متغير معين غير حساس للتغيرات التي تحدث مع مرور الزمن لا يؤدي وظيفته المقصودة.

كما تناول «تراوب» (Traub (1994 أيضاً مسألة ثبات درجات التغيير أو الفرق. ويقترح أن الافتراضات حول المقاييس المعنية قد تكون إشكالية (معضلة) بشكل خاص عندما يخطط المرء للنظر في درجات التغيير الخام (على سبيل المثال، درجات الاختبار البعدي مطروحاً منها درجة الاختبار القبلي أو التمهيدي). وعلى وجه التحديد، يلاحظ «تراوب» أن اتساق ما يقيسه المقياس داخل الفترات الزمنية وعبرها يكتسب أهمية خاصة عند إدراج وتضمين درجات التغيير. على سبيل المثال، قد يمثل الخروج عن القياس المسافي أو الفترتي الحقيقي، حتى خلال فترة زمنية

واحدة، مشكلة لأن التغير في بعض أجزاء من نطاق الدرجات قد لا تكون قابلة للمقارنة من حيث الحجم بالنسبة إلى التغيرات المتطابقة عددياً في نقاط مختلفة على طول المقياس. وبالتالي، فإن التغير من درجة (٢) إلى (٣) قد لا يعادل التغير من درجة (٥) إلى (٦) على مقياس الإعاقة البدنية أو العجز البدني الذي يتراوح بين درجة واحدة إلى عشر درجات. وثمة مثال آخر يتعلق بالتغيرات من فترة زمنية إلى أخرى. قد تتغير الاختلافات (الفروق) فيما يقيسه المقياس نفسه مرتين، إذا استخدم الأشخاص، على سبيل المثال، ذاكرة إجاباتهم السابقة كدليل لتحديد إجاباتهم على الاختبار البعدي بدلاً من أو بالإضافة إلى السمة التي يراد أن يقيسها المقياس. كما يلاحظ «تراوب» أيضاً أن المحاولات الرامية إلى تقليل هذه المشكلة إلى أدنى حد باستخدام مقاييس مختلفة (مثل النماذج المتكافئة أو المتوازية) في الوقت الأول والوقت الثاني قد تؤدي إلى حدوث مشكلات إضافية. واحدة من هذه المشكلات هي المعايير المختلفة أو التدرج المختلف. ما لم تكن وحدات التغير على المقياس الأول تعني بالضبط نفس الشيء الذي تعنيه وحدات التغير في المقياس اللاحق، يمكن أن تكون درجات التغير مضللة. حتى في حالة استخدام الاختبارات المتكافئة المزعومة، يحذر «تراوب» من وجود اختلافات طفيفة في المعايير أو في المتغيرات الكامنة التي تحدد درجاتها. وبالتالي، فإنه ينصح بعدم استخدام درجات التغير.

ويصف «هيويت، وأندرسن، ومينور» (Hewett, Anderson, and Minor (1992) خمس إستراتيجيات تحليلية بديله لتصاميم البحوث التي تنطوي على التغير. بالإضافة إلى (أ) استخدام درجات التغير كمتغيرات تابعة، تتضمن هذه الطرق (ب) اختبار التفاعل بين عضوية المجموعة والوقت، (ج) اختبار الاختلافات أو الفروق في المتوسطات في الوقت الثاني، (د) اختبار الاختلافات أو الفروق في متوسطات الوقت الثاني بعد التحكم في الدرجات في الوقت الأول عن طريق تحليل التباين المصاحب أو المشترك (ANCOVA)، و(هـ) باستخدام تحليل التباين المصاحب أو المشترك ANCOVA اللاعلمي nonparametric للضبط والتحكم في متوسطات الوقت الأول. ويخلص المؤلفون إلى أنه من الضروري مراعاة الدرجات الأولية إذا كانت هذه الدرجات مرتبطة بدرجات لاحقة (كما هو الحال غالباً). وبالتالي، فإنهم يوصون باستخدام حلول تحليل التباين المصاحب أو المشترك ANCOVA. والمنطق هو أن تأخذ في الاعتبار الحالة الأولية للمتغير التابع قبل فحص تأثير المتغير المستقل. يمكن تنظيم تحليل لهذا النوع بطريقتين متماثلتين على الأقل. أحدهما هي التعامل مع المستوى الأولي للمتغير التابع على أنه متغير

مصاحب covariate والتعامل مع مستواه في وقت لاحق كنتيجة. وعلى نحو متساوٍ، يمكن للمرء استبدال درجة التغيير كنتيجة. هاتان الطريقتان متكافئتان من الناحية الرياضية لأنه من بين الدرجات الأولية، والدرجات اللاحقة، ودرجات التغيير، لا يوجد في الواقع سوى جزأين مستقلين من المعلومات. ومعرفة أي اثنين منهما يحدد الثالث (أي، هناك درجتان فقط من الحرية بين الدرجات الثلاث). التحكم في الحالة الأولية للمتغير هو جزء واحد من المعلومات. ويمكن تمثيل الجزء المتبقي باستخدام أي من الجزأين المتبقيين كنتيجة. ومع ذلك، من المهم أن ندرك أن استخدام درجة التغيير كنتيجة عندما يتم التحكم في الحالة الأولية لا يختلف عن استخدام درجة التغيير الخام كنتيجة دون تعديل للحالة الأولية للمتغير. إن إدراج الحالة الأولية كمتغير ضابط أو متغير تحكم يتجنب العديد من المشاكل المرتبطة بعدم ثبات درجة التغيير بسبب أنها درجة التغيير الخام (أي غير المعدلة) التي تخضع لعدم الثبات.

وفي الآونة الأخيرة، لاحظ «كروين وإيمونز وسيجتسما» Kruyen, Emons, and (2014) Sijtsma مشكلات أو قضايا إضافية يمكن أن تصيب أو تكون مصدر إزعاج لدرجات التغيير. وبشكل أكثر تحديدًا، تتناول ورقتهم الاستخدام الشائع لدرجات التغيير، أي تقييم التغيير الفردي في السياق السريري. على سبيل المثال، قد يرغب أحد ممارسي الصحة العقلية في تقييم ما إذا كان المريض قد عانى من أعراض منخفضة أو أقل بعد تلقي العلاج أو ما إذا كان هذا المريض قد تحول وانتقل عبر حد التشخيص وأنه لم يعد من الممكن اعتباره مبررًا لتشخيص سابق. قد لا يتم استخدام ثبات المقياس على نحو مناسب لتحديد ما إذا كانت الاستنتاجات على المستوى الفردي مبررة أم لا. وكما يلاحظ «كروين وآخرون» (Kruyen et al, 2014) أن الثبات هو في الأساس خاصية لأداء المقياس عبر مجموعة من المجيبين. وبالتالي، فإن تطبيقه لإصدار أحكام بشأن التغييرات في الأفراد يثير مشاكل. ينشأ أحد السيناريوهات الشائعة ولكن إشكالي بشكل خاص عندما يحاول الطبيب تحديد مقدار تغير الفرد بناءً على اختلافات الدرجات بمرور الزمن باستخدام مقياس قصير نسبيًا. ويحذر كروين وآخرون (2014) من أنه في الحالات التي يكون فيها المقياس مكوناً من أقل من عشرة بنود، فإن استخدامه بهذه الطريقة يمكن أن يؤدي إلى معدل مرتفع بشكل غير مقبول من الأحكام الخاطئة.

وكما هو الحال بالنسبة للعديد من الموضوعات الأخرى في هذا الكتاب، فإن أفضل الممارسات التي ينبغي اتباعها هي التفكير في الظروف المتاحة وكيفية تعاملها مع البيانات المتاحة والافتراضات الخاصة بمختلف النهج أو الطرق. ستكون هناك ظروف تجعل درجات التغيير ذات مغزى كبير، ولا سيما عندما تؤخذ الدرجات الأولية في الاعتبار، تعمل بشكل جيد. وفي حالات أخرى، ولا سيما عند التفكير في استخدام درجات التغيير الخام، قد يكون هناك ما يبرر اتباع نهج آخر. إن المجموعة الكاملة من الظروف الممكنة والنهج الممكنة كثيرة للغاية بحيث لا يمكن مناقشتها هنا. ومع ذلك، لا ينبغي تجاهل الرسالة الأساسية لهذا القسم - وهي أن ثبات درجة التغيير لا يمكن اعتبارها مساوية لثبات الدرجات لمقياس يتم تطبيقه في وقت واحد.

### الثبات والقوة الإحصائية:

من الفوائد التي كثيراً ما يتم تجاهلها أو التغاضي عنها في كثير من الأحيان في المقاييس الأكثر ثباتاً أنها تزيد من القوة الإحصائية لحجم عينة معين (أو تسمح لحجم عينة أصغر بتوليد قوة مكافئة أو مساوية لحجم عينة أكبر)، مقارنةً بالمقاييس الأقل ثباتاً. وللحصول على درجة محددة من الثقة في القدرة على اكتشاف فرق بمقدار معين أو معنوي بين مجموعتين تجريبيتين، على سبيل المثال، يحتاج المرء إلى عينة ذات حجم معين. ويمكن زيادة احتمال اكتشاف مثل هذا الاختلاف أو الفرق (أي قوة الاختبار الإحصائي) عن طريق زيادة حجم العينة. وفي العديد من التطبيقات، يمكن الحصول على نفس التأثير عن طريق تحسين ثبات المقياس. والمقياس الذي يتمتع بالثبات، مثل العينة الكبيرة، يسهم في تقليل الخطأ نسبياً في التحليل الإحصائي. قد يكون من الأفضل للباحثين أن يقوموا بتقييم المزايا النسبية لزيادة ثبات المقياس مقابل حجم العينة في حالات البحث التي يتوفر فيها كلا الخيارين.

تعتمد القوة المكتسبة من تحسين الثبات على عدد من العوامل، بما في ذلك حجم العينة الأولية، ومستوى الاحتمال المحدد لاكتشاف خطأ من النوع الأول Type I error، وحجم التأثير (مثل، متوسط الفرق) الذي يعتبر معنوياً أو ذا دلالة إحصائية، ونسبة تباين الخطأ الذي يُعزى إلى قياس عدم الثبات بدلاً من عدم تجانس العينة أو مصادر أخرى. وتتطلب المقارنة الدقيقة بين تحسين الثبات وزيادة حجم العينة تحديد هذه العوامل؛ ومع ذلك، توضح الأمثلة التالية هذه النقطة. في حالة بحث افتراضية يكون فيها احتمال وجود خطأ من النوع الأول (0.01)، والفرق يبلغ عشر نقاط بين



المتوسطين يعتبر مهماً، وتباين الخطأ يساوي 100، فإنه يجب زيادة حجم العينة من 128 إلى 172 (زيادة العينة بنسبة 34 %) لرفع قوة اختبار (ف) F من 80.0 إلى 90.0 وسيكون لتقليل تباين الخطأ الإجمالي من 100 إلى 75 (انخفاض بنسبة 25%) نفس النتيجة بشكل أساسي دون زيادة حجم العينة. وقد يحقق استبدال مقياس أقل ثباتاً بمقياس آخر يتمتع بمستوى ثبات عالٍ نفس النتيجة. وكمثال آخر، بالنسبة لعينة حجمها يساوي 50 ( $N = 50$ )، هناك مقياسان بمستوى ثبات قدره 38.0 ويرتبطان معاً بمعامل ارتباط قدره ( $r = 0.24$ ) بالكاد يحققان معنوية أو مستوى دلالة إحصائية عند ( $p < .10$ ). إذا تمت زيادة مستوى ثباتهما إلى 90.0.

إذا بقي الثبات عند 38.0، فستكون هناك حاجة إلى مضاعفة حجم العينة من أجل الوصول إلى معامل ارتباط ذي دلالة إحصائية عند ( $p < .01$ ). ويوفر ليبسي (1990) Lipsey مناقشة أكثر شمولاً حول القوة الإحصائية، بما في ذلك تأثيرات ثبات القياس.

يتم تعزيز وتحسين الثبات إما من خلال عدد البنود أو الفقرات ومتوسط الارتباط بين البنود (والذي ينشأ من ارتباط أقوى بين كل بند والدرجة الحقيقية)؛ إما إضافة المزيد من البنود أو الحصول على بنود أفضل يزيد من الثبات للمقياس. ويمكن لأي من طريقتي تحسين الثبات أن تزيد وتعزز القوة الإحصائية. ومرة أخرى، هذا يشبه ويمثل الوسيلة التي يمكن بها للمعينة (طرق أخذ العينات) أن تزيد من القوة الإحصائية. ويمكن أن يؤدي تقليل خطأ المعينة، إما عن طريق اختيار عينات «أكبر من حيث الحجم» (أي، جمع البيانات من عينة أكبر) أو أخذ عينات «أكثر ذكاءً» (أي، جمع البيانات من الأفراد الذين يمثلون المجتمع الإحصائي موضع الاهتمام أفضل تمثيل)، إلى زيادة القوة الإحصائية. وبالمثل، يمكن أن تقلل المقاييس الأكبر (التي تتضمن بنوداً أكثر) والأكثر ذكاءً (وجود ارتباطات أقوى بين البنود وبعضها البعض) من التباين الناتج عن خطأ القياس، وبالتالي، زيادة القوة الإحصائية.

#### نظرية التعميم:

حتى الآن، لقد ركزت مناقشتنا حول الثبات على تقسيم التباين الملاحظ إلى الجزء الذي يُعزى إلى الدرجة الحقيقية للمتغير الكامن والجزء المتبقي، وهو الخطأ. يقدم هذا القسم بإيجاز إطار عمل أكثر عمومية لتقسيم التباين بين مصادر الخطأ والمصادر الأخرى غير الخطأ. ويعود هذا

الإطار إلى مناقشتنا السابقة للأنهج أو المداخل القائمة على تحليل التباين ANOVA للتمييز بين مصادر التباين.

قبل أن نطبق فكرة التقسيم الأكثر دقة لتباين الخطأ على القياس، دعونا ننظر في مثال بحثي أكثر عمومية يتم فيه فحص مصادر التباين المتعددة. لنفترض أن الباحث أراد تحديد فعالية برنامج تدريبي يهدف إلى زيادة الإنتاجية المهنية. علاوة على ذلك، نفترض أن الباحث قام بتطبيق وإدارة البرنامج التدريبي لعينة كبيرة من أساتذة الجامعات وعلى عينة مماثلة من الفنانين. وحدد الباحث أيضاً مجموعات مماثلة من الأساتذة والفنانين الذين لن يشاركوا في البرنامج التدريبي ولكنهم سيشاركون في نفس تقييم الإنتاجية الذي يشارك فيه المشاركون في البرنامج التدريبي. عند التفكير في الدراسة بعض الشيء، قد يكون هذا الباحث خُص إلى أن ملاحظات الإنتاجية ستعكس تشغيل ثلاثة مصادر محددة للتباين المنهجي أو المنتظم: (١) مشارك مقابل غير مشارك، (٢) أستاذ جامعي مقابل فنان، و(٣) تفاعل هذه الآثار أو التأثيرات. وتتمثل إحدى الإستراتيجيات التحليلية المعقولة أو المنطقية في هذا الموقف في إجراء تحليل التباين ANOVA على درجات الإنتاجية، ومعالجة كل مصدر من مصادر التباين هذه كبعد في التحليل. ويمكن للباحث عندئذ أن يحدد إلى أي مدى يساهم كل مصدر من مصادر التباين في التباين الكلي في الإنتاجية المهنية. في جوهرها، فإن هذه الإستراتيجية التحليلية ستقسم التباين الكلي بين درجات الإنتاجية الملاحظة إلى عدة مصادر هي: المشاركة في التدريب، والمهنة، وتفاعل هذه العوامل، والخطأ. ويمثل الخطأ جميع مصادر التباين بخلاف تلك المحددة بواسطة العوامل السابقة.

الآن، لنأخذ في الاعتبار حالة افتراضية يقوم فيها الباحث بتطوير مقياس الرغبة في الاستقلال الذاتي. وسوف يستخدم هذا المقياس في دراسة للمسنين، والذين قد يواجه بعضهم مشاكل بصرية. وبالتالي، يخطط الباحث لإدارة أو تطبيق مقياس الرغبة في الاستقلال الذاتي شفوياً لهؤلاء الأشخاص الذين سيصعب عليهم القراءة وفي شكل مكتوب للمشاركين المتبقين في الدراسة.

إذا تجاهل الباحث طريقة إدارة أو تطبيق المقياس (المكتوبة مقابل الشفهية) كمصدر للتباين في درجات الاختبار، فإنه سيكون فيما يتعلق بكل درجة تم الحصول عليها بسبب المستوى الحقيقي لرغبة المجيب في الاستقلال الذاتي بالإضافة إلى درجة ما من الخطأ. ويمكن للباحث أن يواصل حساب الثبات كما سبق مناقشته. ومع ذلك، لاحظ أن مجرد حساب معامل "ألفا" على درجات

المقياس دون الأخذ في الاعتبار طريقة إدارة أو تطبيق المقياس لن يميز الخطأ المنهجي المحتمل الناتج عن طريقة تطبيق المقياس عن أي مصدر خطأ آخر.

وكبديل لذلك، يمكن للباحث الاعتراف بوضع طريقة إدارة أو تطبيق المقياس كمصدر للتباين بين الدرجات، باستخدام أسلوب تحليل التباين. إذا أظهرت نتائج التحليل أن الفرق بين أساليب إدارة وتطبيق المقياس يمثل نسبة غير مهمة من التباين الكلي في الدرجات، حينئذ يمكن للباحث أن يكون لديه ثقة أكبر في إمكانية مقارنة الدرجات بالنسبة للأفراد الذين أكملوا الإجابة على المقياس إما بالصيغة الشفوية أو المكتوبة. أما إذا كان هناك مقدار كبير من التباين الكلي الملاحظ في الدرجات يُعزى إلى طريقة إدارة وتطبيق المقياس، فإن الباحث سيعلم أن أي تفسير للدرجات ينبغي أن يأخذ هذا الفرق بين الأنماط في الاعتبار.

توفر نظرية التعميم (على سبيل المثال، Cronbach, Gleser, Nanda, & Rajaratnam, 1972) إطارًا لدراسة مدى إمكانية افتراض تكافؤ عملية القياس عبر بُعد واحد أو أكثر. وفي المثال السابق، كان البعد موضع الاهتمام هو طريقة إدارة وتطبيق المقياس. وكل بُعد من الأبعاد موضع الاهتمام هو مصدر محتمل للتباين ويشار إليه على أنه وجه واحد أو جانب واحد facet. ويركز المثال على أسلوب إدارة وتطبيق المقياس باعتباره المصدر الوحيد المحتمل للتباين (بخلاف الأفراد) الذي يرغب الباحث في تعميمه. لذلك، هذا المثال ينطوي على وجه أو جانب واحد.

وفي لغة نظرية التعميم، تشكل الملاحظات التي يمكن الحصول عليها عبر جميع مستويات أحد الأوجه (على سبيل المثال، مع كل من أسلوب الإدارة أو التطبيق الشفوي والمكتوب للمقياس) عالمًا من الملاحظات المقبولة. ويشار إلى المتوسط لحسابي لهذه الملاحظات باسم الدرجة الكلية universe score وهي تشبه الدرجة الحقيقية في نظرية الاختبار الكلاسيكية (Allen & Yen, 1979) classical test theory. وتسمى الدراسة التي تهدف إلى تحديد مدى قابلية مقارنة الدرجات عبر المستويات المختلفة من الوجه باسم “دراسة إمكانية التعميم generalizability study” أو «G-study». وتعد الدراسة الافتراضية للرغبة في الاستقلال الذاتي مثالاً على دراسة “إمكانية التعميم” بحكم معالجتها لتأثيرات المستويات المختلفة لوجه أو جانب طريقة إدارة أو تطبيق المقياس.

والغرض من دراسة إمكانية التعميم «G-study» هو مساعدة الباحث في تحديد المدى الذي يقوم به الوجه facet في الحد من التعميم أو عدمه. إذا كان هناك وجه (على سبيل المثال، أسلوب الإدارة أو التطبيق) يفسر قدرًا كبيرًا من التباين في الدرجات الملاحظة، فإن النتائج لا تعمم عبر المستويات (مثل الإدارة الشفوية مقابل الإدارة المكتوبة) لهذا الوجه أو الجانب. ويتم التعبير عن مدى إمكانية التعميم عبر مستويات الوجه facet دون تحريف البيانات بمعامل إمكانية التعميم generalizability coefficient. ويتم حساب هذا المعامل عادةً عن طريق تكوين نسبة من متوسط المربعات المناسبة الناتجة عن تحليل التباين التي يتم إجراؤها كجزء من دراسة إمكانية التعميم G-study. ومن الناحية النظرية، فإن معامل إمكانية التعميم هو نسبة التباين في الدرجات الكلية إلى تباين الدرجات الملاحظة وبالتالي فهو مماثل ومشابه لمعامل الثبات (Allen & Yen, 1979). ومع ذلك، لاحظ أنه إذا أسفرت دراسة إمكانية التعميم عن معامل تعميم ضعيف، فإن تصميم الدراسة يشير إلى مصدر المشكلة (أي، الجانب أو الوجه الذي تم فحصه). فمعامل الثبات لا يحدد سوى مقدار الخطأ دون أن يعزى ذلك إلى مصدر محدد.

وفي بعض الحالات، قد يكون اختيار تصميم تحليل التباين المناسب، وتحديد التأثيرات التي تتوافق أو تتطابق مع جوانب الاهتمام، وبناء معامل إمكانية التعميم الصحيح أمراً صعباً. كما هو الحال مع تحليل التباين بشكل عام، يمكن لأبعاد متعددة، وتأثيرات متداخلة، ومتقاطعة، ومختلطة أن تجعل دراسة إمكانية التعميم معقدة. (انظر مايرز ١٩٧٩؛ أو كيرك ١٩٩٥، Myers 1979, or Kirk 1995، للاطلاع على المناقشات العامة حول تصميمات تحليل التباين) وينصح بالحفاظ على تصميم دراسة إمكانية التعميم بسيطة. ومن الحكمة أيضاً الرجوع إلى مصدر يشرح بالتفصيل كيفية بناء نموذج تحليل التباين المناسب لنوع معين من دراسات إمكانية التعميم. ويصف «كروكر والجينا» (Crocker and Algina, 1986) التصميمات المناسبة للعديد من دراسات مختلفة لها جانب أو وجه واحد واثنين من دراسات إمكانية التعميم. كما يوفر هذا المصدر مقدمة عامة جيدة لنظرية التعميم.

#### ملخص:

تتمتع المقاييس بالثبات بقدر ما تتكون من بنود أو فقرات ثابتة تشترك في متغير كامن عام أو مشترك. ويتوافق معامل «ألفا» بشكل وثيق مع التعريف التقليدي للثبات كنسبة التباين في

المقياس الذي يعزى إلى الدرجة الحقيقية للمتغير الكامن. الطرق المختلفة لحساب الثبات لها فائدة مختلفة في حالات معينة. على سبيل المثال، إذا لم يكن لدى الشخص إمكانية الوصول إلى إصدارات متكافئة من مقياس معين، فإنه من المستحيل حساب ثبات النماذج البديلة. الباحث الذي يفهم مزايا وعيوب الأساليب البديلة لحساب الثبات يكون في وضع أفضل لاتخاذ أحكام مستنيرة عند تصميم دراسة القياس أو تقييم تقرير (دراسة) منشور. إن الثبات هو قضية تتجاوز الاهتمامات والسياقات البحثية المختلفة. وتستخدم كلا من نظرية إمكانية التعميم ونهج معامل الارتباط داخل الطبقة لتقييم الثبات وكلا الطريقتين تستغل قدرة تحليل التباين لعزل مصادر التباين المنفصلة وقياسها كمياً. وهكذا، فإن الأنشطة البحثية التي تبدو على أنها متباينة كدراسة لإمكانية التعميم، والتقييم النفسي (السيكومتري) لبنود الاستبانة، وتقييم ثبات المقيمين. جميعها تقوم على تعريف عام ومشارك للثبات كنسبة التباين التي تعزى إلى الدرجة الحقيقية للظاهرة موضع الاهتمام.

#### تمارين:

- ١- إذا كانت هناك مجموعة من البنود أو الفقرات لديها اتساق داخلي جيد، ماذا يعني ذلك بشأن علاقة هذه البنود بالمتغير الكامن الخاص بها؟
- ٢- في هذا التمرين ٣، افترض أن ما يلي هو مصفوفة تباين لمقياس (Y) يتكون من ثلاثة بنود هي (X1 و X2 و X3):

١,٢	٠,٥٠	٠,٤٠
٠,٥٠	١,٠	٠,٦٠
٠,٤٠	٠,٦٠	١,٨

أ - ما هو التباين لكل من البنود الثلاثة (X1 و X2 و X3)؟

ب- ما هو تباين المقياس (Y)؟

ج- ما قيمة معامل «ألفا» للمقياس (Y)؟

- ٣- ناقش الطرق التي يربك بها أو يخلط أسلوب ثبات إعادة الاختبار test-retest reliability عوامل أخرى بخصائص المقياس الفعلي.

٤- كيف ينبثق منطق ثبات النماذج البديلة من افتراضات الاختبارات المتكافئة؟

٥- عند فحص زوج من الدرجات على نفس المتغير لنفس الأفراد، تم تطبيقه في وقتين مختلفين (على سبيل المثال، درجة الاختبار القبلي تسمى  $X1$  ودرجة الاختبار البعدي تسمى  $X2$ )، كيف يمكن لقوة الارتباط بين الدرجات (أي الارتباط الذاتي الخاص بهم) أن تؤثر على ثبات درجة التغير المحسوبة بطرح درجة من الأخرى. على سبيل المثال،  $(X1 - X2)$ ؟

٦- تمثل جميع معاملات الثبات تقريباً نسبة من نوع واحد من المعلومات بالنسبة لنوع آخر. في الواقع، هذه النسبة تشكل أساساً تعريف الثبات. اشرح باختصار ما هما نوعا المعلومات الأساسيان لتعريف الثبات وأنه، عند التعبير عنهما كنسبة، هما الأساس لمعاملات الثبات.

٧- ما هي بعض الانتقادات لاستخدام معامل ألفا؟ ومتى تنطبق هذه الانتقادات؟ وما هي العلاجات أو البدائل الموجودة للاستخدام التقليدي للمعامل ألفا، ومتى تكون مناسبة؟

ملاحظات:

١- بالنسبة للبنود الموزونة، يتم ضرب التغيرات في حاصل ضرب التباينات في مربعات أوزان البنود المقابلة لها. انظر (Nunnally, 1978, pp. 154-156) للحصول على وصف أكثر وضوحاً.

٢- خلال الكتاب، سوف يظهر حل أي تمرين يتطلب إجابة رقمية في حاشية.

٣- الإجابات: (أ) ١,٢، ١,٠، ١,٨، (التي تبلغ 0.4)؛ (ب) ٧,٠ (مجموع جميع البنود في المصفوفة)؛ (ج)  $(2/3) * [(0.7/0.4)-1] = 64.0$ .

## الفصل الرابع

### الصدق

في حين يتعلق الثبات بمدى تأثير متغير على مجموعة من البنود، فإن الصدق يتعلق بما إذا كان المتغير هو السبب الأساسي لتغاير البنود. وبقدر ما يكون المقياس ثابتاً، فإن التباين في درجات المقياس يمكن أن يُعزى إلى الدرجة الحقيقية لبعض الظواهر التي تمارس تأثيراً سببياً على جميع البنود أو الفقرات. ومع ذلك، فإن تحديد درجة ثبات المقياس لا يضمن أن المتغير الكامن الذي تتشارك البنود فيه، في الواقع، هو المتغير المهم لمطور المقياس. وأن مدى كفاية المقياس كمقياس لمتغير محدد (مثل الضغوط النفسية المدركة) هي مسألة صدق.

وقد أعطى بعض المؤلفين معنى أوسع للصدق. على سبيل المثال، وصف ميسيك Messick (1995) ستة أنواع أو أنماط من الصدق، أحدها (الصدق التتابعي validity consequential) يتعلق بتأثير الطريقة التي يتم بها استخدام درجاتهم على المجيبين. وعلى الرغم من أن آراء “ميسيك” بشأن الصدق أثارت بعض القضايا المثيرة للتفكير، فإن نظام التصنيف الخاص به لم يُعتمد على نطاق واسع. ووفقاً للتفسير الأكثر تقليدية (المحافظ)، يتم استنتاج الصدق من الطريقة التي تم بها بناء وتكوين المقياس، أو قدرته على التنبؤ بأحداث محددة، أو علاقته بمقاييس مفاهيم بنائية أخرى. وهناك ثلاثة أنواع أساسية من الصدق التي تتوافق مع هذه العمليات وهي:

1- صدق المحتوى Content validity.

2- الصدق المرتبط بمحك أو بمعيار Criterion-related validity.

### 3- الصدق البنائي أو صدق التكوين الفرضي Construct validity.

وسيتم استعراض كل نوع بإيجاز. وللحصول على معالجة أكثر شمولاً للصدق، بما في ذلك مناقشة القضايا المنهجية والإحصائية المتعلقة بالصدق المرتبط بمحك ومؤشرات الصدق البديلة، انظر الفصل العاشر في كتاب “غيسيلي، وكامبل، وزيدك” Zedeck and ,Campbell ,Ghiselli (1981). وقد يرغب القراء أيضاً في النظر في عرض ميسيك (1995) الأكثر شمولاً لموضوع الصدق.

#### صدق المحتوى أو المضمون:

يتعلق صدق المحتوى بكفاية عينة البنود - أي المدى التي تعكس فيها مجموعة محددة من البنود نطاق أو مجال المحتوى. ويكون تقييم صدق المحتوى أسهل عندما يكون المجال أو النطاق (على سبيل المثال، جميع مفردات الكلمات التي يتم تدريسها لطلاب الصف السادس) محدداً بشكل جيد. وتكون المسألة أو المشكلة أكثر صعوبة عند قياس السمات مثل المعتقدات أو المواقف والاتجاهات أو التصرفات لأنه من الصعب تحديد بالضبط ما هو نطاق البنود المحتملة ومتى تكون عينة من البنود ممثلة للنطاق أو المجال. من الناحية النظرية، يتمتع المقياس بصدق المحتوى عندما تكون بنوده مجموعة فرعية تم اختيارها عشوائياً من مجموعة كلية من البنود المناسبة. في مثال اختبار المفردات المستخدم أعلاه (مفردات الكلمات التي يتم تدريسها لطلاب الصف السادس)، يمكن تحقيق ذلك بسهولة. حيث سيتم تعريف كل الكلمات التي تم تدريسها خلال العام الدراسي على أنها المجموعة الكلية للبنود. وبعد ذلك يمكن أخذ مجموعة فرعية من هذه الكلمات كعينة. ومع ذلك، في حالة قياس المعتقدات، على سبيل المثال، لا توجد لدينا قائمة ملائمة من البنود ذات الصلة، إلا أن أحد أساليب أو طرق تطوير مقياس ما (على سبيل المثال، مراجعة البنود من قبل الخبراء من حيث صلتها بمجال الاهتمام، كما هو مقترح في الفصل الخامس) يمكن أن تساعد في زيادة ملائمة البند. على سبيل المثال، إذا كان الباحث بحاجة إلى وضع مقياس يتناقض مع النتائج المتوقعة والنتائج المرجوة (على سبيل المثال، توقع إشراك المريض في صنع القرار مقابل رغبة الطبيب في إشراك المريض في صنع القرار)، فقد يكون من المستحسن إثبات أن جميع النتائج ذات الصلة كانت ممثلة في البنود. وللقيام بذلك، قد يكون للباحث زملاء على دراية بسياق البحث يقومون بمراجعة قائمة



أولية من البنود واقتراح مجالات المحتوى التي تم إهمالها ولكن ينبغي إدراجها في المقياس. ويمكن بعد ذلك إضافة بنود تعكس هذا المحتوى.

يرتبط صدق المحتوى ارتباطاً وثيقاً بتعريف المفهوم البنائي قيد الدراسة. وبكل بساطة، يجب أن يعكس محتوى المقياس التعريف المفاهيمي أو التصوري المنطبق على هذا المقياس. قد تكون بعض المفاهيم قد تم تعريفها بأكثر من طريقة من قبل المنظرين أو قد تقع عند تقاطع أو التقاء مفاهيم متعددة. ومن الضروري أن يتضمن محتوى البنود جوانب الظاهرة التي تم توضيحها في تعريفها المفاهيمي أو التصوري وليست الجوانب الأخرى التي قد تكون ذات صلة ولكنها خارجة عما يقصده الباحث لتلك الأداة المعنية.

وكمثال على ذلك، قام «ستربا وزملاؤه» (Sterba and colleagues (2007 بتطوير مقياس الفعالية الزوجية المتعلقة بكيفية إدراك الأزواج الذين أصيب أحدهم بالتهاب المفاصل الروماتويدي لقدرتهم على إدارة المرض كفريق واحد. وكانت هذه الأداة هي الأولى لتقييم الفعالية الزوجية، وثقة الزوجين في قدرتهم كفريق واحد على اتخاذ مختلف الإجراءات لتعزيز الصحة. وهكذا، كان المفهوم الأساسي للأداة مختلفاً عن المفاهيم الأخرى مثل الفعالية الذاتية الفردية، على الرغم من أنه ذو صلة بها. وكانت المرحلة الأولى من هذا الجهد عبارة عن دراسة لتطوير البنود والتي شكلت الأساس لادعاءات المؤلفين حول صدق المحتوى. وتهدف هذه الدراسة إلى تحديد المحتوى المناسب من الأدبيات التجريبية والنظرية الأوسع نطاقاً لإدراجها في المقياس. على الرغم من أن المؤلفين قاموا بفحص المحتوى لمقاييس المفاهيم البنائية ذات الصلة (على سبيل المثال، الفعالية الذاتية الخاصة بالتهاب المفاصل)، فقد قاموا بتوجيه تطوير البنود الخاص بهم إلى سمات أو خصائص محددة للمفهوم البنائي كما حدوده. وتنطوي أحد الجوانب المهمة من هذا التعريف على مدى شعور الأزواج بالثقة بأنهم، كفريق واحد، يمكنهم إدارة تحديات المرض. وبناءً على ذلك، فقد اشتملت دراسة تطوير البنود على مقابلات مع الأزواج لمعرفة أفكارهم حول ما إذا كان المفهوم البنائي يتوافق معهم، ولفهم كيفية تصورهم له، وتحديد اللغة التي استخدموها لوصفه. إن الأفكار والرؤى المستقاة من التعريف النظري للمفهوم البنائي، ومراجعة الأدبيات، والمقابلات مع المرضى تم الاسترشاد بها عند بناء وتكوين البنود. وقد تمت كتابة البنود بشكل صريح للتعبير عن جانب تصور فريق الأزواج فيما يتعلق بالفعالية. وقد تضمنت هذه العملية أن محتوى البنود يعكس المفهوم البنائي المحدد الذي كان الباحثون مهتمين به وليس مفاهيم أخرى مختلفة (مثل الفعالية الذاتية أو

الدعم الزوجي المدرك) والتي قد تكون ذات صلة بها من الناحية النظرية. وكخطوة إضافية للتحقق من صدق المحتوى، فقد طلب «ستربا وآخرون»، (Sterba et al. (2007 من مجموعة من خبراء المحتوى مراجعة البنود التي طورها فريق البحث في البداية في ضوء تعريفهم النظري. وكان هذا الإجراء بمثابة تحقق إضافي من أن البنود كانت تمثل المحتوى ذا الصلة الذي تم تصميم الأداة لقياسه. وإجمالاً، فقد زادت هذه الخطوات من احتمال تضمين وإدراج المحتوى ذي الصلة في المقياس في حين لم يتم إدراج المحتوى غير ذي الصلة، مما يدعم الادعاءات بصدق المحتوى.

#### نطاق المتغير والآثار المترتبة على صدق المحتوى:

ثمة مسألة أو قضية مرتبطة ارتباطاً وثيقاً بصدق المحتوى تتمثل في نطاق scope المتغير موضع الاهتمام وبالتالي البنود التي تُولف المقياس. من المحتمل أن يكون الباحث مهتماً بمتغير شامل مثل نوعية الحياة أو تفضيلات الشراء أو بمتغير أكثر تحديداً (ضيقاً) مثل السعادة أو الوعي بالتكلفة. وبطبيعة الحال، فإن المحدد الأساسي لما نقيسه ينبغي أن تكون الأسئلة التي نسعى للإجابة عليها وفهمنا للقضايا النظرية ذات الصلة بتلك الأسئلة. وفي الوقت نفسه، ينبغي أن نضع في اعتبارنا مدى صلة اتساع المفهوم البنائي الخاص بنا بصدق المحتوى.

ومن الواضح، فإن المحتوى المبالغ فيه أكثر من اللازم أو المحتوى المختصر الأقل من اللازم يمكن أن يثير مشاكل إذا أدى إلى وجود عدم تطابق بين المفهوم الذي يراد من المقياس قياسه ونطاق البنود التي يحتوي عليها. في سياق الاختبارات التعليمية، حذر “ميسيك” Messick (1995) من المخاطر المحتملة لـ “سوء أو نقص تمثيل المفهوم”، أي أخذ عينات محدودة جداً من مجال ونطاق المحتوى، وبالتالي من المحتمل أن يحرم أحد الممتحنين من فرصة إثبات وإظهار معرفته أو معرفتها. ومع ذلك، يشير أيضاً إلى خطر “التباين غير الملائم للمفهوم البنائي” الذي يمكن أن ينتج عن أخذ عينات من المحتوى على نطاق واسع جداً. على الرغم من أن “ميسيك” يناقش هاتين الظاهرتين على أنهما تهديد للصدق البنائي، إلا أنهما تتعلقان بوضوح باتساع نطاق اختيار المحتوى وبالتالي صدق المحتوى.

وعلى الرغم من أن إستراتيجية اختيار عينة المحتوى والتي تكون إما مفرطة في الشمول أو شديدة التقييد يمكن أن تمثل إشكالية، ويبدو أن هناك تفاوتاً أو عدم تناسق، مع وجود مفاهيم بنائية أوسع نطاقاً ومجموعات بنود أكثر شمولاً. ومن مظاهر هذا الاتجاه أنه يبدو، في بعض السياقات

البحثية، أن هناك تفضيلاً متزايداً لمقاييس القياس ذات الأهداف العريضة أو التي تكون على نطاق واسع. والقياس المتعلق بالحالة الصحية هو مثال على ذلك. ومن الواضح أن هناك بعض الأسباب التي تقف وراء هذا التفضيل المتزايد للمقاييس الأكثر شمولاً. إذا كان المقياس يمكن أن يغطي سمة واسعة نسبياً، فمن المحتمل أن يكون أكثر فائدة واستخداماً. ويمكن لشخص أن يقول، على سبيل المثال، بأن مقياساً عاماً للإعاقة قد يكون أكثر جاذبية من مقياس خاص بمجموعة معينة من المرضى. وهذا لا يغني عن الحاجة إلى وضع وتطوير مقياس جديد لكل نوع من الأمراض المختلفة فحسب، بل إنه يعزز المقارنات بين المجموعات لأنه يمكن قياس كليهما باستخدام نفس الأداة. وعلاوة على ذلك، فإن أدوات القياس الأكثر قابلية للتطبيق يمكن أن تؤدي مهمة أو وظيفة مزدوجة، وهي تلبية الاحتياجات الإدارية فضلاً عن الاحتياجات البحثية. ولأغراض مثل تلبية ومعالجة حقوق ومطالبات الإعاقة، على سبيل المثال، يمكن تطبيق أداة أكثر عمومية بشكل موحد لقياس الإعاقة عبر الظروف (الحالات) ومن ثم تحديد الأهلية للحصول على تعويض بطريقة موحدة بين المرضى والسياقات (الحالات). كذلك، قد تستفيد مقارنات فعالية العلاج عندما يتم قياس نتائج العلاج بنفس الأداة. ومع ذلك، قد تخفي هذه المزايا الظاهرة أهمية الاختلافات السياقية والعينات المحددة في طريقة أداء المقياس.

وكما هو الحال بالنسبة للثبات، فإن الصدق ليس خاصية متأصلة في أداة القياس بل هو خاصية لأداة القياس في سياق استخدامها. قد تكون الأداة صادقة في سياق واحد ولكنها غير صادقة في سياق آخر أو عند استخدامها بشكل مختلف. كمثال واضح للغاية أو بديهي على ذلك، فإن جهاز ضغط الدم في حالة عمل جيدة هو وسيلة صادقة لتقييم الضغط في الأوعية الدموية للشخص؛ ومع ذلك، إذا تم وضع الجهاز حول رأس شخص ما في محاولة لقياس ذكائه، فإن النتيجة لن تكون صحيحة أو صادقة. لا يوجد الصدق في الأداة نفسها ولكن في كيفية وطريقة استخدامها. وينطبق هذا أيضاً على المقاييس في القياس القائم على البنود. ولذلك، فإن صدق محتوى أداة ما سيعتمد أو يتوقف على مدى جودة وسلامة البنود التي تتألف منها خريطة المقياس للمفهوم البنائي موضع الاهتمام للمجتمع الإحصائي وسياق البحث المحدد. وسيؤثر تنوع محتوى البند على خريطة العمل هذه.

يتمثل أحد أوجه القصور المحتملة في الأداة التي تحاول قياس مفهوم ما على نطاق واسع في أن البنود التي تم اختيارها لتمثيل وتجسيد هذا المفهوم قد لا تنطبق بالتساوي في جميع المواقف

والحالات أو على كل المجتمع الإحصائي. في الواقع، قد يكون صدق المحتوى للأداة معتمداً إلى حد كبير على السياق. لنفكر، على سبيل المثال، أداة افتراضية تهدف إلى تقييم (قياس) الوعي بالتكاليف. ولأغراض تدريبية أو تعليمية، دعونا نفترض أن الأعمال السابقة قد كشفت عن أن بعض الأفراد مدركون بدرجة عالية للتكلفة عند إجراء عمليات الشراء. سوف يمتنع هؤلاء الأفراد عن إجراء عملية الشراء، حتى لو كانت السلع المعروضة جذابة للغاية، إذا كانوا يعتقدون أن التكلفة لا تمثل قيمة استثنائية. ولنفترض كذلك أن اشخاصاً آخرين كان وعيهم أقل بمتغير التكلفة. بالنسبة لهؤلاء الأفراد، فإن التكلفة ليست عاملاً محدداً رئيسياً في قراراتهم المتعلقة بالشراء، ولكنها ليست سوى واحدة من عدة اعتبارات مهمة تشمل الجودة والرغبة. وإذا كانت سلعة معينة من السلع الاستهلاكية جذابة للغاية بالنسبة لهم، فقد يفضلون الحصول عليها عندما تتاح الفرصة نفسها بدلاً من انتظار إمكانية العثور على سعر أقل لنفس السلعة أو المماثلة لها في أماكن أخرى.

ومع الاستمرار في هذا المثال الافتراضي، دعونا نفترض أن الباحث يريد تكوين مقياس لتقييم الوعي بالتكاليف على النحو المبين هنا. قد يحقق ذلك من خلال التفكير بعناية حول ما يشمل المفهوم ثم تطوير البنود التي تعكس ذلك التعريف العملي للمفهوم.

ومن المسائل أو القضايا التي سيواجهها الباحث مدى اتساع أو ضيق نطاق قرارات الشراء التي ينبغي أن تتضمنها وتشملها الأداة. وهذا، بدوره، ينبغي أن يتم الاسترشاد به في تكوين وبناء البنود. وفي إحدى الحالات المتطرفة، يمكن تركيز البنود بشكل محكم إلى حد ما على خيارات شراء محددة، مثل شراء سيارة جديدة. وعلى النقيض من ذلك، يمكن للباحث أن يحاول وضع مقياس عام يمكن تطبيقه على أي نوع من الحالات التي قد تكون فيها التكلفة قضية أو مشكلة، بما في ذلك مجموعة كبيرة من السلع والخدمات. دعونا ننظر في بعض الأمثلة على أنواع البنود التي قد يسفر عنها كل نهج من هذين النهجين. افترض أن مجموعتي البنود التاليتين طلبتا من المجيبين الموافقة على كل بند على مقياس اتفاق مكون من ست نقاط مع خيارات استجابة تتراوح بين لا أوافق بشدة إلى أوافق بشدة. البنود أرقام 1 و2 و3 التالية تخص الوعي بالتكلفة في سياق شراء سيارة.

١ - أفضل تأجيل الحصول على السيارة التي أريدها حقاً بدلاً من دفع بنس واحد زيادة لذلك.

٢- أفضل أن أدفع أكثر قليلاً وأحصل على السيارة التي أريدها بالضبط بدلاً من توفير بعض المال ولكن يجب المساومة.

٣- أنا أعتبر السعر الأول الذي يذكره مندوب مبيعات السيارات ليس أكثر من حيلة مساومة. وتتضمن المجموعة الثانية من البنود (أي من البند رقم 4 إلى البند رقم 6) نسخة عامة من المجموعة الأولى، وصيغت لتكون قابلة للتطبيق عبر السياقات المختلفة.

٤- أنا أفضل تأجيل امتلاك ما أريده حقاً بدلاً من دفع بنس واحد أكثر لذلك مما يجب عليّ دفعه.

٥- أفضل أن أدفع أكثر قليلاً وأحصل على السلع أو الخدمات التي أريدها بالضبط بدلاً من توفير بعض المال ولكن يجب المساومة.

٦- أنا أعتبر أن السعر الأول المذكور لأي سلع أو خدمات ليس أكثر من حيلة مساومة.

ويبدو أن البنود الثلاثة الأولى تستفيد من الوعي بالتكلفة وتنتشر في وجهة نظر معينة حول أسعار السيارات وأهميتها في اتخاذ خيارات شراء السيارات. قد يوافق بعض الناس بشدة أو لا يوافقون على العبارات بصيغتها المكتوبة، ولكن ليس من الصعب أن نتصور أن الاستجابات التي اختارها شخص ما ستكون متشابهة إلى حد ما عبر مختلف حالات شراء السيارات. كما يبدو من المعقول أن تنطبق هذه البنود بشكل أو بآخر بنفس الطريقة تقريباً على غالبية الأشخاص الذين يواجهون قرار شراء سيارة. وعلى الرغم من أن مدى تأييد مختلف الناس للفقرات من شأنه أن يختلف على الأرجح اختلافاً كبيراً، إلا أنني أتوقع درجة عالية نسبياً من الاتساق في مستويات الموافقة التي يقدمها أي شخص.

دعونا الآن ننتقل إلى العبارات الثلاث المتبقية، أي البنود من رقم 4 إلى رقم 6. وليس من الصعب تصور أن مجال المتغير الأوسع نطاقاً (أي النطاق الشامل، وليس المحدد بالسياق للوعي بالتكلفة) وما يرتبط به من بنود قد يسبب بعض الصعوبات. ولا أتوقع أن يتفق معظم الناس على هذه البنود على نحو مماثل عبر سياقات مختلفة بشكل ملحوظ. على سبيل المثال، ينبغي النظر في كيفية استجابة الناس على هذه البنود في ثلاث حالات أو مواقف مختلفة: (أ) شراء تلفزيون، (ب)

شراء عمل أو فن نادر، (ج) تقرير ما إذا كان ينبغي تلقي المشورة القانونية أو التخلي عنها في مسألة ذات أهمية حيوية. وإنني أشك بشدة في أن هذه السياقات المختلفة ستحسن علاقة هذه البنود مع متغيرها الكامن المقصود منها، وأن معرفة كيفية استجابة شخص ما في أحد تلك السياقات لن تكون بالضرورة تنبؤية لكيفية استجابته للسياقين الآخرين. وأساس هذا الشك هو اعتقادي بأن الوعي بالتكلفة كما تعكسه هذه البنود الأخيرة له معنى مختلف اختلافاً جوهرياً في سياقات مختلفة. وأهميته بالنسبة للمتغيرات الأخرى التي تؤثر على الاستجابات للبنود تكون غير متسقة. ويبدو أن الوعي بالتكلفة الذي يحفز المساومة (التفاوض) على تكلفة المشورة القانونية التي قد تؤدي إلى بعض العواقب الخطيرة مثل السجن أو الخسارة الكاملة لأصول المرء هو متغير مختلف تماماً عن المتغير الذي ينطوي عليه التفاوض والمساومة حول سعر تلفزيون جديد. وإذا أكدت الأدلة التجريبية هذه الشكوك، فإنني سأقول إن المجموعة الثانية من البنود الأكثر عمومية لم تكن ذات مضمون صحيح عبر السياقات. وعلاوة على ذلك، أود أن أقترح أن عدم صدق الأدلة في سياقات معينة (مثل تحديد الخيارات المتعلقة بالحصول على مساعدة قانونية أو طبية هامة) ينشأ، على الأقل جزئياً، عن عدم خصوصية وتحديد المفهوم البنائي. ونتيجة لهذا النقص في التحديد، قد يتعرض صدق محتوى البنود من البند رقم 4 إلى البند رقم 6 للخطر.

وبطبيعة الحال، يمكن للمرء أن يجادل بأن البنود من البند الأول إلى البند الثالث ستكون أيضاً محتوى غير صالح في مواقف أو حالات أخرى غير شراء سيارة. وعلى الرغم من أن هذا صحيح بلا شك، إلا أن الفرق المهم هو أن القيد في صدق محتوى تلك البنود واضح، وأن الأشخاص العقلانيين لن يحاولوا استخدامها لتقييم الوعي بالتكلفة في سياقات مختلفة بشكل ملحوظ. ومع ذلك، في حالة وجود البنود الأكثر عمومية، فإن عدم ملاءمتها لبعض المواقف يكون أقل وضوحاً ويمكن بسهولة ألا يلاحظه أحد.

في المثال السابق، كان صدق المحتوى معرضاً للخطر عندما تم استخدام نسخ أكثر عمومية من البنود لأن مفهوم الوعي بالتكلفة يأخذ في الأساس معنى مختلفاً في بعض السياقات. وثمة طريقة أخرى يتم من خلالها تعريف أوسع نطاقاً للمفهوم البنائي ربما يعرض صدق المحتوى للخطر وهي عندما تكون مجموعة فرعية من البنود ذات صلة بالمفهوم البنائي أحياناً وأحياناً لا تكون كذلك. وهذا يعني، أنه في بعض الظروف، ستكون بعض البنود مصدراً لحدوث تباين غير ذي صلة بالمفهوم البنائي (ميسك، 1995، Messick)، حتى عندما تحتفظ بنود أخرى بصلتها وأهميتها.

وكمثال افتراضي آخر، ضع في اعتبارك مجموعة فرعية من البنود قيد النظر لمقياس عام للإعاقة. (مرة أخرى، يمكننا أن نفترض أن خيارات الاستجابة المصاحبة تتراوح من لا أوافق بشدة إلى أوافق بشدة).

أ - لدي مشكلة في استخدام يدي للمهام التي تتطلب قوة قبضة.

ب- لدي مشكلة في الوقوف من وضعية الجلوس.

ج - عندما أبقى في وضع واحد لفترة طويلة جداً، فإن أطرافي تصبح صلبة متيبسة.

د - أنا أتعب بسهولة عندما أبذل جهداً كبيراً جداً.

هـ - لم يعد بإمكانني القيام بأشياء كان بإمكانني القيام بها بسهولة.

يمثل كل بند من هذه البنود قيداً وظيفياً يمكن أن يعتبر ممثلاً للإعاقة بصورة منطقية. وبالنسبة لبعض أنواع المرض، قد يتوقع أن تكون كلها مظاهر للحالة وقد تميل إلى أن تحدث في انسجام. غير أن هناك مشكلة تنشأ إذا كان الغرض من البنود هو تقييم الإعاقة بالنسبة لحالة لا تتضمن كل هذه القيود. وفي مثل هذه الحالة، قد تكون بعض البنود تتمتع بصدق المحتوى في حين أن البعض الآخر ليس كذلك. لذلك، على سبيل المثال، بالنسبة لشخص يعاني من كسر حاد في الساق، فإن البند رقم «أ» والذي يصف عدم وجود قوة القبضة سيكون بنداً لا يتمتع بصدق المحتوى. أي إن قوة القبضة ستكون غير مفيدة فيما يتعلق بتأثر كسر ساق الشخص. في المقابل، بالنسبة لبعض الحالات الأخرى، ربما السكتة الدماغية أو التهاب المفاصل الروماتويدي، فإن بنداً عن قوة القبضة قد يتمتع بصدق المحتوى. لذلك، في حين أن إدراج وتضمين البند رقم «أ» يضر بصدق المحتوى لمن يعاني من ساق المكسورة، فإن حذفه أو إهماله قد يضعف صدق المحتوى للمريض الذي يعاني من التهاب المفاصل الروماتويدي. وتكمن الصعوبة هنا في أن التعريف الإجرائي الضمني للإعاقة الذي اعتمدته عند تكوين هذه البنود التوضيحية كان فضفاضاً وواسعاً أكثر مما ينبغي. ونتيجة لذلك، وضعت قائمة بمظاهر الإعاقة المقبولة التي قد تظهر في مجموعة من الحالات الحادة أو المزمنة دون مراعاة للكيفية التي يمكن أن تناسب هذه القائمة، أي حالة محددة تؤدي إلى الإعاقة. ومن ثم، فإن التعريف الأوسع نطاقاً يؤدي مرة أخرى إلى نشوء حالة يمكن أن تضر بصدق المحتوى.

تم وضع الأمثلة السابقة لتوضيح وإثارة نقطة. وإنني أشك في أن أي شخص سيفكر فعلياً في وضع وتطوير مقياس للوعي بالتكلفة مع فكرة أنه يمكن استخدام بنود مثل البنود من البند الرابع إلى البند السادس في أي سياق دون عقاب (مخاطرة). كما أن أي باحث خبير (ذي خبرة) لن يجمع بين البنود من البند الأول إلى البند الخامس في مقياس واحد يهدف إلى تقييم الإعاقة عبر مجموعة واسعة من الظروف والحالات. ومع ذلك، توضح هذه الأمثلة المتطرفة نقطة تنطبق في بعض حالات القياس في العالم الحقيقي أو من واقع الحياة: عندما يكون المتغير والمقياس المراد استخدامه ذو نطاق واسع للغاية، يزداد احتمال عدم تطبيق بعض المؤشرات (وبالتالي لن يكون له صدق محتوى) في حالات معينة أو لمجموعات معينة.

ومن النتائج السلبية الأخرى المحتملة لتعريف متغير ما على نطاق واسع أن خطوات إثبات صدق المحتوى قد يصبح أكثر عبئاً وإرهاقاً. فعلى سبيل المثال، كما هو موضح في الفصل الخامس، من الشائع أن يطلب من الخبراء مراجعة محتوى البنود كوسيلة لضمان صدق المحتوى. ومع اتساع نطاق المفهوم البنائي، قد تحتاج أيضاً الخبرة الجماعية للأشخاص الذين طُلب منهم مراجعة صدق المحتوى إلى التوسع. كما أن المهمة المعروضة على هؤلاء الخبراء قد تصبح أكثر صعوبة لأن هؤلاء المحكمين يضطرون إلى النظر في متى قد يكون بند معين ذا صلة بالمفهوم البنائي أو لا يكون له صلة. وهذا يتطلب أن يتصوروا أو يتخيلوا شيئاً ما يقترب من جميع الحالات والمواقف الممكنة التي يمكن فيها استخدام الأداة وتقييم مدى ملاءمته لكل منهما - وهو مهمة وعمل هائل. ومع تناقص خصوصية المفهوم البنائي بدقة، قد يجد المحكمون صعوبة أكبر في تقييم ما إذا كانت هناك بنود محددة تناسب المفهوم البنائي على نحو كافٍ. وقد يتطلب ذلك إنشاء فريق من المحكمين الخبراء أكبر مما يلزم مع تحديد أكثر دقة للمفهوم البنائي لتحقيق نفس الدرجة من الثقة في تقديراتهم وتقييماتهم.

وعندما تتعلق أسئلة البحث ذات الاهتمام بمتغيرات واسعة النطاق بطبيعتها، قد يفكر الباحثون فيما إذا كان مقياس عام واحد أو مقاييس محددة متعددة تكون أكثر ملاءمة للمهمة وأكثر ملاءمة لإثبات صدق المحتوى. ومواصلةً لمثال الوعي بالتكاليف، قد يفكر الباحث فيما إذا كانت عدة مقاييس خاصة بسياقاتها الفريدة من نوعها ستكون بدلاً أكثر فائدة لمجموعة غير محددة من البنود المماثلة لمجموعة البنود من البند الرابع إلى البند السادس السابق ذكرها. أو قد يكون من



المفيد، في حالة مقياس الإعاقة، اختيار مجموعات البنود التي تشير تحديداً إلى أنواع القيود التي يمكن توقعها بين مجموعة محددة بشكل ضيق للغاية من الناس.

يختلف صدق المحتوى عن الأنواع الأخرى من الصدق التي سنناقشها لأنه يُعرّف فعلياً بالإجراءات التي يتخذها منشئ أو مبتكر المقياس في بداية تطوير المقياس. وأهم هذه العوامل هو النظر بعناية في المتغير الذي يقصد من المقياس أن يقيسه ومن ثم أنواع البنود التي تكون عينات كافية من المجال أو النطاق الذي يتضمنه المتغير. وتستحق المفاضلات بين النطاقات الأكثر عمومية والنطاقات الأكثر تحديداً ومجموعات البنود اهتماماً دقيقاً في هذه المرحلة من عملية تطوير المقياس ويمكن أن يكون لها تأثير كبير على صدق محتوى البنود التي تم اختيارها في نهاية المطاف. وبالمثل، ينبغي للباحث، عند اختيار مقياس موجود بالفعل لاستخدامه في دراسة ما، بدلاً من مجرد الاعتماد على التقارير السابقة المتعلقة بصدق المحتوى من الدراسات السابقة، أن يفكر بعناية في كيفية توافق أو تطابق محتوى المقاييس المتاحة مع المجتمع الإحصائي، والسياق، وفحص أسئلة الدراسة المحددة التي تم التخطيط لها.

ونظراً لأن اتساع المتغير الذي يختار الباحث دراسته يحدد بطريقة مباشرة وشفافة حجم المجموعة المحتملة من المحتوى الذي يجب أن يمثلها مطور المقياس بشكل مناسب، فقد فضلت مناقشة هذه القضية جنباً إلى جنب مع صدق المحتوى. ومع ذلك، فإن اتساع النطاق الذي يأمل مقياس ما أن يشمل، تترتب عليه أيضاً آثار للأنواع الأخرى من الصدق التي أناقشها في الجزء المتبقي من هذا الفصل. ولن أكرر ما قمت بتغطيته حتى الآن، ولكن يجب أن يدرك القراء أن ملائمة ومناسبة البنود للمفهوم البنائي أمر بالغ الأهمية بالنسبة لجميع أنواع الصدق التي تم استعراضها في هذا الفصل. ومع اتساع حدود هذا المفهوم البنائي، يمكن أن تصبح مهمة إيجاد وإثبات الصدق المرتبط بالمحك (بالمعايير) أو الصدق البنائي أكثر صعوبة، تماماً كما هو الحال بالنسبة لصدق المحتوى.

الصدق المرتبط بمحك (معياري):

من أجل الحصول على الصدق المرتبط بمحك أو بمعيار، كما يوحي المصطلح، لا يشترط إلا أن يكون لبند أو مقياس ارتباط تجريبي ببعض المحكات أو “المعيار الذهبي” المفترض. وسواء قد تم فهم الأساس النظري لذلك الارتباط أم لا، فإن ذلك ليس له صلة بالصدق المرتبط بالمعيار.

على سبيل المثال، إذا تبين أن عملية التغطيس (البحث بعصا الاستنباء عن الماء) مرتبطة بشكل تجريبي بتحديد موقع مصادر المياه الجوفية، فإن عملية التغطيس ستكون صادقة فيما يتعلق بمعيار حفر الآبار بنجاح. وبالتالي، فإن الصدق المرتبط بالمحك في حد ذاته هو مسألة عملية (تطبيقية) أكثر من كونها مسألة علمية، لأنها ليست معنية بفهم العملية بل مجرد التنبؤ بها وتوقعها. وفي الواقع، غالباً ما يشار إلى الصدق المرتبط بمحك على أنه صدق تنبؤي predictive validity.

لا يعني الصدق المرتبط بمعيار تحت أي اسم بالضرورة وجود علاقة سببية بين المتغيرات، حتى عندما يكون الترتيب الزمني للتنبؤ والمعيار واضح لا لبس فيه. وبطبيعة الحال، قد يكون التنبؤ في سياق النظرية (مثل التنبؤ كفرضية) ذا صلة بالعلاقات السببية بين المتغيرات ويمكن أن يخدم أو يؤدي غرضاً علمياً مفيداً.

وثمة نقطة أخرى جديرة بالملاحظة بشأن الصدق المرتبط بمعيار، هي أن المرء، من الناحية المنطقية، يتناول نفس نوع قضية الصدق سواء كان المعيار يتبع أو يسبق أو يتزامن مع القياس المعني. وهكذا، بالإضافة إلى الصدق التنبؤي، فإن الصدق المتزامن أو التلازمي concurrent validity (على سبيل المثال، التنبؤ بمهارة القيادة من الإجابات على الأسئلة الشفوية التي طرحت في اختبار القيادة) أو حتى الصدق اللاحق postdictive validity (على سبيل المثال، التنبؤ بالوزن عند الولادة من مقياس حالة النمو للطفل الرضيع) يمكن استخدامها بشكل مترادف إلى حد ما مع الصدق المرتبط بمحك أو معيار. وأهم جانب من جوانب الصدق المرتبط بمحك أو معيار ليس العلاقة الزمنية بين المقياس المعني (الذي يراد التثبت من صدقه) والمعيار الذي يحاول المرء استنتاجه من قيمته، بل هو قوة العلاقة التجريبية بين الحدثين. ويتمتع الصدق المرتبط بمحك بميزة مقارنة بالمصطلحات الأخرى تتمثل في كونه محايداً زمنياً، وبالتالي فهو الأفضل.

#### الصدق المرتبط بمحك مقابل الدقة:

وقبل الانتهاء من مناقشة الصدق المرتبط بمحك، هناك بضع كلمات تتعلق بعلاقته بالدقة. وكما يشير «غيسيلي وآخرون (1981) Ghiselli et al.»، فإن معامل الارتباط، الذي كان المؤشر التقليدي للصدق المرتبط بمحك، قد لا يكون مفيداً للغاية عندما تكون الدقة التنبؤية هي القضية أو المسألة. فعلى سبيل المثال، لا يكشف معامل الارتباط عن عدد الحالات التي يتم تصنيفها بشكل

صحيح من قبل أحد المؤشرات (على الرغم من أن الجداول التي تقدم تقديرًا لنسبة الحالات التي تقع في فئات مئوية مختلفة، استنادًا إلى حجم الارتباط بين المنبئ والمعيار، تم وصفها من قبل "غيسيلي وآخرون"، ص 311). وقد يكون من الأنسب في بعض الحالات تقسيم كل من المتنبئ ومعياره إلى فئات منفصلة وتقييم "معدل الدخول" لحالات الاستبدال إلى الفئة الصحيحة للمعيار استنادًا إلى فئة التنبؤ الخاصة بهم. فعلى سبيل المثال، يمكن تصنيف كل متغير إلى فئات «منخفضة» مقابل «عالية» كما يمكن تصور الدقة على أنها نسبة التصنيفات الصحيحة (أي الحالات التي تتطابق فيها قيمة المؤشر مع قيمة المعيار). حيث إن تقسيم الشخص للفئات هو اعتبار مهم. لنأخذ في الاعتبار معياراً أو محكاً له حالتان غير تعسفتين، مثل «مريض» و«صحيح أو معافى»، وأداة تقييم تحتوي على مجموعة من الدرجات التي يريد المحقق تقسيمها. والغرض من أداة التقييم هو التنبؤ بما إذا كان الأشخاص سيختبرون على أنهم إما موجب أو سالب بالنسبة للمرض موضع الاهتمام. ونظرًا لأن النتيجة ستكون ثنائية الاستجابة، فمن المنطقي أن يكون المتنبئ ثنائي. هناك خطآن محتملان في التصنيف هما: يمكن أن يصنف المقياس عن طريق الخطأ شخصًا مريضًا حقًا على أنه صحيح (سليم) (نتيجة سلبية زائفة أو خاطئة) أو شخصًا صحيحًا حقًا على أنه مريض (نتيجة إيجابية زائفة أو خاطئة).

حيث يمكن أن يؤثر التقسيم على معدلات هذين النوعين من الأخطاء على امتداد نطاق الدرجات من أداة التقييم. في النهايات، سيؤدي تصنيف كل شخص على أنه صحيح إلى تجنب أي سلبات خاطئة أو زائفة (ولكن يزيد من الإيجابيات الزائفة) بينما يؤدي تصنيف كل شخص على أنه مريض إلى تجنب أي إيجابيات زائفة (ولكن يزيد من السلبات الزائفة). ومن الواضح، في كلتا الحالتين المتطرفتين، أن أداة التقييم لن تكون لها أي قيمة تنبؤية على الإطلاق. والهدف، بالطبع، هو اختيار نقطة القطع التي ينتج عنها أقل عدد من الأخطاء من أي نوع، وبالتالي أعلى درجة دقة. وفي كثير من الأحيان، لا توجد نقطة قطع مثالية (أي، نقطة قطع واحدة تؤدي إلى تصنيف مثالي). وفي مثل هذه الحالة، قد يبذل الباحث مجهودًا واعياً لتقليل نوع واحد من الأخطاء إلى الحد الأدنى بدلاً من الآخر. فعلى سبيل المثال، إذا كان المرض مدمراً وكان العلاج فعالاً وغير مكلف وغير مضر، فإن تكلفة نتيجة سلبية زائفة (تؤدي إلى التفريط في العلاج) أكبر بكثير من تكلفة نتيجة إيجابية زائفة (تؤدي إلى الإفراط في العلاج). وبالتالي، فإن اختيار نقطة قطع لتقليل السلبات

الزائفة مع قبول إجابيات زائفة يبدو مناسباً. ومن ناحية أخرى، إذا كان العلاج مكلفاً ومزعجاً «كريهاً» في آن واحد وإذا كان المرض خفيفاً أو معتدلاً، فقد تكون المفاضلة العكسية أكثر منطقية.

أيضاً، من المهم أن نتذكر أنه حتى لو كان الارتباط بين مقياس التنبؤ والمعيار تاماً، فإن النتيجة التي تم الحصول عليها في مقياس التنبؤ ليست تقديراً للمعيار أو المحك. معاملات الارتباط غير حساسة للتحويلات الخطية لأحد المتغيرين أو كليهما. ويشير الارتباط المرتفع بين متغيرين على أن الدرجات على هذين المتغيرين والتي تم الحصول عليها من نفس الشخص سوف تشغل مواقع مماثلة في توزيعات كل منهما. على سبيل المثال، شخص ما يسجل درجات عالية جداً على المتغير الأول من المحتمل أيضاً أن يسجل درجات عالية جداً على المتغير الثاني إذا كان الاثنان مرتبطين بشدة. ومع ذلك، فإن التعبير «عالية جداً» هو مصطلح نسبي وليس مصطلحاً مطلقاً ولا يأخذ في الاعتبار وحدتي القياس الخاصة بالمتغيرين، على سبيل المثال. وقد يكون من الضروري تحويل وحدات قياس مقياس التنبؤ إلى وحدات القياس الخاصة بالمعيار للحصول على تنبؤ رقمي أو كمي دقيق. وهذا التعديل يعادل أو يساوي تحديد القاطع intercept المناسب بالإضافة إلى ميل خط الانحدار. ويمكن أن يؤدي عدم إدراك الحاجة إلى تحويل درجة ما إلى استنتاجات خاطئة. ومن المحتمل حدوث خطأ من هذا النوع على الأرجح إذا حدث أن تمت معايرة مقياس التنبؤ في وحدات تقع في نفس نطاق المعيار. افترض، على سبيل المثال، أن شخصاً ما وضع «مقياس مخالفة السرعة» التالي وذلك للتنبؤ بعدد تذاكر مخالفة السرعة التي سيحصل عليها السائقون على مدى خمس سنوات:

١- أنا أتجاوز الحد الأقصى للسرعة عندما أقود السيارة.

(كثيراً، أحياناً، نادراً، أبداً).

٢- على الطرق متعددة المسارات، وأنا أقود سيارتي أتجاوز الممر العابر.

(كثيراً، أحياناً، نادراً، أبداً).

٣- أنا أحكم بنفسي على ما هي سرعة القيادة المناسبة.

(كثيراً، أحياناً، نادراً، أبداً).

ولنفترض أيضاً افتراضاً غير منطقي بأن المقياس يرتبط ارتباطاً تاماً بعدد تذاكر مخالفة السرعة التي تم الحصول عليها خلال فترة خمس سنوات. يتم تسجيل المقياس عن طريق إعطاء كل بند القيمة (3) عندما يختار المجيب الاستجابة «كثيراً» والقيمة (2) للاستجابة «أحياناً» والقيمة (1) للاستجابة «نادرًا» والقيمة (صفر) للاستجابة «أبدًا». ثم يتم جمع درجات البنود للحصول على درجة المقياس. درجة صدق المعيار الكاملة لا تعني أن الدرجة البالغة تسع درجات تشير إلى الحصول على تسع تذاكر مخالفة للسرعة على مدار خمس سنوات، بل يعني هذا أن الأشخاص الذين يسجلون أعلى الدرجات على الأداة هم أيضاً الأشخاص الذين لديهم أكبر عدد من تذاكر مخالفة السرعة في السنة. ومن شأن بعض عمليات التحويل المحددة تجريبياً (مثلاً،  $0.33 * \text{الدرجة}$ ) أن تسفر عن التقدير الفعلي. من شأن هذا التحويل المحدد أن ينتجاً بثلاث بطاقات للسائق الذي حصل على تسع درجات. وكلما زاد الصدق المرتبط بمعيار، كان التقدير أكثر دقة استناداً إلى مقياس التنبؤ. ومع ذلك، فإن التشابه بين القيم العددية للمعيار ومقياس التنبؤ قبل القيام بعملية تحويل مناسبة لن تكون له أي علاقة بدرجة الصدق.

#### الصدق البنائي (صدق التكوين الفرضي):

إن الصدق البنائي كما أشار كرونباخ وميهل (1955) Cronbach & Meehl يُعنى مباشرةً بالعلاقة النظرية لمتغير (على سبيل المثال، درجة على مقياس ما) بمتغيرات أخرى. وهو المدى الذي يتصرف (يسلك) فيه المقياس بالطريقة التي يفترض أن يسلكها المفهوم البنائي الذي يستهدف قياسه فيما يتعلق ببناء المقاييس للمفاهيم البنائية الأخرى.

لذلك على سبيل المثال، إذا نظرنا إلى متغير معين، استناداً إلى النظرية، على أنه مرتبط إيجاباً بالمفهومين البنائيين A و B، ويرتبط سلباً بالمفهومين البنائيين C و D، وغير مرتبط بالمفهومين البنائيين X و Y، فإن المقياس الذي يهدف إلى قياس ذلك المفهوم البنائي يجب أن تكون له علاقة مماثلة لمقاييس تلك المفاهيم البنائية الأخرى. وبعبارة أخرى، يجب أن يرتبط مقياسنا بشكل إيجابي مع مقاييس المفهومين البنائيين A و B، ومرتبطة سلباً بمقاييس المفهومين البنائيين C و D، وغير مرتبط بمقاييس المفهومين البنائيين X و Y. ويمكن التعبير عن هذه العلاقات الافتراضية كما في المخطط في الشكل رقم (1-4).

ويوفر مدى تطابق الارتباطات التجريبية مع النموذج المتوقع بعض الأدلة على مدى تصرف أو أداء المقياس كما يفعل المتغير المفترض أن يقيسه.

وبتطبيق هذا النهج العام، وصف “ستيربا وآخرون” (2007) Sterba et al. نمطاً من الارتباطات التي تنبأوا بأن تكون لأداتهم مع عدة مفاهيم بنائية ذات صلة. على سبيل المثال، استناداً إلى تحليلهم النظري، توقعوا أن ترتبط درجات الفعالية الزوجية مع مقاييس الجودة الزوجية، والتكيف النفسي، ومعايير العمل الجماعي. وقد حددت بعض هذه التوقعات وجود ارتباطات سلبية، مثل العلاقة العكسية المفترضة بين الفعالية الزوجية والاكنتاب. وكان أحد القيود على تلك الدراسة، التي أقرها المؤلفون، أن الفعالية الذاتية الفردية لم يتم تقييمها. إن إثبات أن الفعالية الزوجية تختلف عن تصورات الكفاءة الذاتية الفردية للشركاء داخل الأزواج couples يُعد جانباً مهماً محتملاً في عملية التحقق من صدق المقياس.

	A	B	C	D	X	Y
Variable	+	+	-	-	0	0

شكل (1-4): علاقة افتراضية بين متغيرات

التمييز بين الصدق البنائي والصدق المرتبط بمحك:

غالبًا ما يخلط الناس بين الصدق البنائي والصدق المرتبط بمحك أو معيار وذلك بسبب أن الارتباط الدقيق نفسه يمكن أن يخدم كلا النوعين من الصدق. ويمكن الفرق في مقصد وهدف الباحث أكثر مما يكمن في القيمة التي تم الحصول عليها. على سبيل المثال، قد يحاول أخصائي الأوبئة تحديد أي من مجموعة متنوعة من المقاييس التي تم الحصول عليها في دراسة مسحية ترتبط بالحالة أو الوضع الصحي. وقد يكون القصد أو الهدف من ذلك هو مجرد تحديد عوامل الخطر دون أي قلق (على الأقل في البداية) للآليات السببية الأساسية التي تربط بين الدرجات على المقاييس والحالة الصحية. الصدق، في هذه الحالة، هي الدرجة التي يمكن أن تتنبأ بها المقاييس بالحالة الصحية. وكبديل لذلك، يمكن أن يكون الاهتمام هو النظرية والتفسير بشكل أكبر. وقد يؤيد الباحث، مثله مثل أخصائي الأوبئة الذي تم وصفه في الفصل الافتتاحي من هذا الكتاب، نموذجاً نظرياً يعتبر الإجهاد سبباً في الحالة الصحية، وقد تكون المسألة هي مدى نجاح مقياس تم تطويره حديثاً في قياس الإجهاد. قد يتم تقييم ذلك من خلال تقييم سلوك المقياس بالنسبة للكيفية التي تشير

النظرية في الكيفية التي يجب أن يعمل بها الإجهاد. إذا كانت النظرية تشير إلى ضرورة الارتباط بين حالة الإجهاد والحالة الصحية، فيمكن استخدام نفس العلاقة التجريبية المستخدمة كدليل على الصدق التنبؤي في المثال السابق كدليل على الصدق البنائي.

وما يسمى بالتحقق من صدق المجموعات المعروفة known-groups validation هو مثال آخر على الإجراء الذي يمكن تصنيفه إما على أنه صدق بنائي أو الصدق المرتبط بمعيار، تبعاً لنية وهدف الباحث. يتضمن التحقق من صدق المجموعات المعروفة عادةً إثبات أن مقياس محدد يمكن أن يميز أعضاء مجموعة واحدة عن أعضاء مجموعة أخرى بناءً على درجات المقياس الخاصة بهم. قد يكون الغرض إما النظرية ذات صلة (كما هو الحال عندما يتم التحقق من قياس الاتجاهات نحو مجموعة معينة عن طريق التمييز بشكل صحيح بين أولئك الذين ينتمون أو لا ينتمون إلى عضوية تلك المجموعة) أو التنبؤية البحتة (كما هو الحال عندما يستخدم المرء مجموعة من البنود ليس لها علاقة على ما يبدو بالتنبؤ بترك الوظيفة). في الحالة الأولى، ينبغي اعتبار الإجراء نوعاً من الصدق البنائي، وفي الحالة الثانية، يكون الصدق هو الصدق المرتبط بمعيار أو محك.

وبالإضافة إلى النية أو الهدف، هناك اختلاف آخر يكمن في كثير من الأحيان وراء الصدق المرتبط بالمعيار والصدق البنائي. غالباً ما يتم تقييم الصدق المرتبط بمعيار مباشرةً عن طريق حساب ارتباط بين المقياس الذي يتم التحقق من صدقه والمعيار (على سبيل المثال، بعض السلوكيات أو الحالة أو الدرجة). وعلى النقيض من ذلك، لا يمكن تقييم الصدق البنائي إلا بصورة غير مباشرة (انظر لورد ونوفيك Lord & Novick, 2008, p. 278)، لأن المقارنة ذات الصلة تكون بمتغير كامن وليس متغير ملاحظ. وفي هذا الصدد، فإنه يشبه الثبات. وفي الفصول السابقة، لاحظنا أن الثبات يتعلق في الحقيقة بالعلاقة بين مؤشر ودرجة حقيقية لا يمكن ملاحظتها، وأنها نستنتج تلك العلاقة على أساس الارتباطات بين المؤشرات التي يمكن ملاحظتها. وينطبق نفس الشيء على الصدق البنائي. حيث لا يمكننا أن نحسب مباشرة الارتباط بين أداة القياس الذي يتم التحقق من صدقها والمتغير الكامن، ولكن يجب علينا أن نفعل ذلك بصورة غير مباشرة من خلال ملاحظة الارتباطات بين أداة القياس الجديدة والمؤشرات الأخرى الموثوق بها للمتغير الكامن. وفي الواقع، قد يجد الباحثون أنه من المفيد استخدام هذا الاختلاف أو الفرق كقاعدة حسابية تقريبية عندما يكونون غير متأكدين مما إذا كان الصدق المرتبط بمعيار أو الصدق

البنائي هو موضع الاهتمام الأساسي. إذا كان الهدف هو التنبؤ بمخرجات أو نتائج يمكن ملاحظتها (على سبيل المثال، السلوك، أو الحالة، أو الدرجة الملاحظة)، فإن الصدق المرتبط بمعيار أو محك قد يكون هو الهدف. ومن ناحية أخرى، إذا كان الهدف هو التنبؤ بمستوى مفهوم بنائي افتراضي، لا يمكن ملاحظته عن طريق مؤشر يمكن ملاحظته، فمن المرجح أن يكون الهدف هو الصدق البنائي.

ومع ذلك، في بعض الأحيان، سيتضمن التحقق من الصدق المرتبط بمعيار مؤشراً يُقصد به أن يعكس أو يظهر بعض الدرجات الحقيقية غير الملاحظة. وهذا هو الحال عندما يكون المعيار (على سبيل المثال، خريج كلية الحقوق المؤهل للقبول في نقابة المحامين) قائماً على شكل ما من أشكال التقييم النفسي (على سبيل المثال، اجتياز امتحان المحاماة). إذا ظل هدف الباحث هو مجرد تنبؤ بدون تفصيل مفاهيمي، فإن مثل هذه المقارنة هي مثال على الصدق المرتبط بمعيار على الرغم من عدم المقارنة بين المتنبئ والنتائج موضع الاهتمام.

#### التخفيف (التضعيف) Attenuation:

إلى الحد أو القدر الذي يكون فيه مؤشران اثنان لا يتمتعان بالثبات التام، فإن أي ارتباط بينهما سوف يقلل من تقدير الارتباط بين الدرجات الحقيقية المقابلة لكل منهما (على سبيل المثال، لورد ونوفيك 2008, Lord & Novick). وعندما نفترض أن الخطأ عشوائي، فإنه لا يمكن الربط إلا بين الأجزاء التي تتمتع بالثبات من كل من المؤشرين فقط. وبناءً على ذلك، يضعف (ينخفض) معامل الصدق الملاحظ المحسوب بالربط بين متغيرين يمكن ملاحظتهما كنتيجة لأي عدم ثبات متأصل في تلك المؤشرات. ويمكن للشخص تطبيق تصحيح الضعف من علاقة ارتباط ملحوظ تأخذ في الاعتبار عدم ثبات المتغيرات. وينطوي هذا التصحيح على تقسيم الارتباط الملاحظ بين المؤشرين على الجذر التربيعي لحاصل ضرب ثبات كل منهما. وهكذا،

$$r(T)_{xy} = \frac{r_{xy}}{\sqrt{r_{xx} r_{yy}}}$$

حيث  $r(T)_{xy}$  هي الارتباط بين الدرجات الحقيقية للمتغيرات  $X$  و  $Y$ ، و  $r_{xy}$  هي العلاقة بين الدرجات الملاحظة، و  $r_{xx}$  و  $r_{yy}$  هي ثبات المتغيرات  $X$  و  $Y$ ، على التوالي.



وعلى الرغم من توافر مثل هذه التصحيحات، فإن بعض المؤلفين أو الكتاب يعارضون هذه التصحيحات. فعلى سبيل المثال، يشير “نونالي وبيرنشتاين” Bernstein and Nunnally (1994) إلى أنها قد تضلل الباحثين إلى الاعتقاد بأن الارتباط أقوى مما هو عليه في الواقع. كما يلاحظون أن المعاملات التي تم تصحيحها يمكن أن تتجاوز أحياناً قيمة الواحد الصحيح (١,٠) (ص 257)، وهذا يمثل مشكلة. ويشير لورد ونوفيك Lord and Novick (2008) إلى أنه عندما يتم التقليل من تقدير ثبات كل من المقياسين (والتي تحدث في مقام معادلة التصحيح)، فإن الارتباط بين الدرجات الحقيقية سيكون مبالغاً في تقديره (ص 138). وبالإشارة إلى ما تم تناوله في الفصول السابقة من أنه وفقاً لنماذج أخرى غير الاختبارات المتكافئة، فإن معامل «ألفا» يكون تقديراً للحد الأدنى للثبات الحقيقي للمقياس، يصبح من الواضح أن ينشأ بسهولة «تصحيح» غير دقيق. وبالتالي، هناك حجج عملية قوية ضد استخدام التصحيح للتضعيف عند فحص الارتباطات، سواء من أجل تقييم الصدق أو لأغراض أخرى. وبسبب الطبيعة غير المباشرة لمعاملات الارتباط المستخدمة في تقييم الصدق البنائي، كما تمت مناقشته في وقت سابق، فقد يكون من المغري بشكل خاص تصحيح معاملات الصدق من أجل التضعيف. ومع ذلك، لا تزال الحجج ضد تعديل التضعيف منطبقة، ومن الممارسات الشائعة عدم تصحيح معامل الارتباط في مثل هذه الحالات.

إلى أي مدى يجب أن تكون الارتباطات قوية لإظهار الصدق البنائي؟

لا توجد نقطة قطع تحدد الصدق البنائي. ومن المهم أن ندرك أن هناك مقياسيين قد يتشاركان في أكثر من مفهوم بنائي متشابه. على وجه التحديد، فإن أوجه التشابه في الطريقة التي يتم بها قياس المفاهيم البنائية قد تفسر بعض التباين في الدرجات المستقلة عن التشابه في المفهوم البنائي. على سبيل المثال، فإن متغيرين اثنين تم تسجيلهما على نظام تسجيل متعدد النقاط (الدرجات من 1 إلى 100) سيكون لهما ارتباط أعلى مع بعضهما البعض مقارنة بمتغير ثنائي، مع تساوي كل شيء آخر. وهذا يكون خادعاً بسبب هيكل أو بناء طرق القياس. وبالمثل، نظراً للتشابه الإجرائي، قد ترتبط البيانات من نوع واحد تم جمعها عن طريق المقابلات بدرجة ما مع البيانات الأخرى التي تم جمعها بنفس الطريقة - أي إن بعض التباين بين متغيرين قد يكون بسبب تشابه القياس بدلاً من تشابه المفهوم البنائي. وتوفر هذه الحقيقة الأساس للإجابة على السؤال المتعلق

بحجم الارتباطات اللازمة لاستنتاج الصدق البنائي. وينبغي أن تظهر المتغيرات، كحد أدنى، التباين أعلى وأكثر مما يمكن أن يعزى إلى تباين الطريقة المشتركة.

#### مصفوفة الصفات المتعددة - الأساليب المتعددة Multitrait- Multimethod matrix:

ابتكر «كامبل وفيسكي» (1959) Campbell and Fiske إجراءً يسمى مصفوفة الصفات المتعددة - الأساليب المتعددة multitrait-multimethod matrix التي تُعد مفيدة للغاية لفحص ودراسة الصدق البنائي. ويتفق هذا النهج أو الطريقة تماماً مع الفكرة، التي تم تقديمها مسبقاً، التي ترى أن الصدق البنائي يتم تقييمه بصورة غير مباشرة ويجب الاستدلال عليه من المؤشرات المتاحة للمتغير الكامن محل الاهتمام. وينطوي الإجراء على قياس أكثر من مفهوم بنائي واحد بأكثر من طريقة بحيث يتم الحصول على مصفوفة «مقاطعة تماماً» حسب مصفوفة القياس. على سبيل المثال، لنفترض أنه قد تم تصميم دراسة ما يتم فيها قياس كل من القلق والاكتئاب ومقاس أو حجم الحذاء في مرتين منفصلتين باستخدام إجراءات قياسيَّين مختلفين في كل مرة. (لاحظ أنه كان من الممكن قياس عينتين مختلفتين من الأفراد في نفس الوقت. ما هو تأثير ذلك على منطق النهج أو الطريقة؟) ويمكن تقييم كل مفهوم بنائي بطريقتين: مقياس تناظري بصري (خط أو سطر يضع عليه المجيبون علامة تشير إلى مقدار السمة التي يمتلكونها، سواء كانت القلق أو الاكتئاب أو كبر حجم القدم) وتصنيف معين من قبل الشخص الذي يجري المقابلة بعد ١٥ دقيقة من التفاعل مع كل موضوع. يمكن للمرء بعد ذلك بناء مصفوفة من الارتباطات التي تم الحصول عليها بين القياسات كما هو مبين في الجدول رقم (٤-١).

هناك تمييز آخر محتمل، غير موجود في الجدول، وهو بين السمات ذات الصلة مقابل السمات غير ذات الصلة. لأن المدخلات «البنود» التي تعكس نفس السمة (المفهوم البنائي) ونفس الطريقة ينبغي أن تشترك في كل من تباين الأسلوب «الطريقة» وتباين المفهوم البنائي، فمن المتوقع أن تكون هذه الارتباطات أعلى. ومن المؤمل أن تكون الارتباطات المقابلة لنفس السمة ولكن الأساليب مختلفة هي التالية في العلو. إذا كان الأمر كذلك، فإن هذا يشير إلى أن التباين للمفهوم البنائي أعلى من التباين للأسلوب أو الطريقة. وبعبارة أخرى، إن مقاييسنا كانت أكثر تأثراً بما تم قياسه من كيفية أو طريقة قياسه. وعلى النقيض من ذلك، لا يوجد سبب يدعو إلى وجود أي تباين بين حجم الحذاء وأي من المفهومين البنائيَّين الآخرين عندما يتم قياسهما بإجراءات مختلفة. ومن

ثم، يجب ألا تختلف هذه الارتباطات اختلافاً كبيراً عن الصفر. بالنسبة للمفاهيم البنائية غير المتطابقة ولكنها مرتبطة من الناحية النظرية، مثل الاكتئاب والقلق، قد يتوقع الشخص وجود بعض التغيرات للمفهوم البنائي.

ومن المحتمل أن تكون هذه مجموعة معلومات مفيدة للغاية من الارتباطات المتبادلة لإيجاد وإثبات الصدق البنائي. على سبيل المثال، إذا كانت مقاييس الاكتئاب لدينا راسخة ومستقرة جيداً على حد سواء ولكن يجري حالياً وضع وتطوير مقاييس القلق لدينا، يمكننا تقييم مقدار التغير الذي يعزى إلى تشابه المفاهيم في ظل إجراءات قياس مماثلة ومختلفة. تؤكد النظرية على أن القلق والاكتئاب ينبغي أن يكونا مرتبطين إلى حد كبير حتى لو تم قياسهما بطرق مختلفة. وإذا ثبت أن هذا هو الحال، فإنه سيكون بمثابة دليل على الصدق البنائي لمقياس القلق الجديد لدينا. وبشكل أكثر تحديداً، فإن هذه الارتباطات تشير إلى الصدق التقاربي convergent validity، أي دليلاً على التشابه بين مقاييس المفاهيم البنائية المرتبطة نظرياً. ومن الناحية المثالية، فإن الارتباطات بين القلق والاكتئاب ستكون أقل من تلك التي توجد بين اثنين من مقاييس الاكتئاب أو اثنين من مقاييس القلق ولكن أكبر بكثير من بين أي من درجات الاكتئاب وحجم الأحذية. وبنفس القدر من الأهمية دليل على أن مقاييس القلق لم ترتبط ارتباطاً ذا دلالة إحصائية بمقاييس حجم الحذاء، بغض النظر عن التشابه أو الاختلاف في تقنية القياس. هذا دليل على الصدق التمايزي validity discriminant (ويسمى أحياناً الصدق المتباين divergent validity)، وعدم وجود ارتباط بين مقاييس المفاهيم البنائية غير ذات الصلة. يشير الارتباط ذي الدلالة الإحصائية بين حجم الحذاء والقلق عند قياسها بنفس الطريقة إلى أن الطريقة في حد ذاتها تمثل مقداراً كبيراً من التباين (والتغير) المرتبط بمقاييس مماثلة للمفاهيم البنائية المختلفة.

جدول (4-1) تفسيرات الارتباطات في مصفوفة الصفات المتعددة - الأساليب المتعددة

		Time 1					
Time 2		$A_v$	$A_i$	$D_v$	$D_i$	$S_v$	$S_i$
	$A_v$	TM	T	M		M	
	$A_i$	T	TM		M		M
	$D_v$	M		TM	T	M	
	$D_i$		M	T	TM		M
	$S_v$	M		M		TM	T
	$S_i$		M		M	T	TM

ملاحظات: (TM) تساوي نفس السمة ونفس الطريقة أو الأسلوب (الثبات)؛ و (T) نفس السمة، وطريقة مختلفة؛ و (A) و (D) و (S) تشير إلى التكوينات الفرضية الفلق، والاكتئاب، وحجم الأحذية على الترتيب. الرموز السفلية (v) و (i) تشير إلى أساليب تناظرية بصرية وطرق المقابلة، على التوالي.

ولاحظ ميتشل (1979) Mitchell أن الأساليب والطرق التي تنطوي عليها عملية جمع البيانات لمصفوفة الصفات المتعددة - الأساليب المتعددة تشكل دراسة إمكانية التعميم ذات وجهين أو جانبين two-facet G-study (أو دراسة قابلية التعميم؛ انظر الفصل الثالث)، مع اعتبار أن السمات والأساليب هي الأوجه أو الجوانب.

وتتيح لنا مصفوفة الصفات المتعددة - الأساليب المتعددة تقسيم التغيرات إلى مصادر «الأسلوب» ومصادر «السمة» (أو المفهوم البنائي). ويمكننا بعد ذلك تقديم بيانات أكثر دقة حول الصدق البنائي، لأنها تسمح لنا بالتمييز بين التغيرات الذي يعكس بالفعل التشابه بين المفهوم البنائي (وبالتالي يرتبط بالصدق البنائي) عن التغيرات الذي هو نتاج تطبيق إجراءات قياس متشابهة (وبالتالي لا يرتبط بالصدق البنائي). ولا يمكن تحقيق هذا التمييز عندما ينظر الشخص ببساطة في ارتباط واحد بين اثنين من المقاييس.

ماذا عن الصدق الظاهري؟

يستخدم العديد من الأشخاص مصطلح «الصدق الظاهري» لوصف مجموعة من البنود التي تقيم ما يبدو أنها تقيسه. في رأيي، هذا الاستخدام سيئ لعدة أسباب:

أولاً، الافتراض بأن المقياس يقوم بتقييم ما يبدو أنه يقيسه قد يكون خاطئاً. فعلى سبيل المثال، قام كل من «إدلر وبنياميني» (Idler and Benyamini (1997 بفحص 27 دراسة وبائية كبيرة وأجريت بطريقة جيدة لتحديد ما هو البند المشترك الذي تمت الاستفادة منه على وجه الدقة. يطلب هذا البند من الناس تقييم صحتهم بشكل عام على أنها إما سيئة أو متوسطة أو جيدة أو ممتازة. معظم الناس سيحكمون على هذا المقياس المكون من بند واحد لتقييم ما يقولونه بالضبط: صحة المستجيبين. وقد لاحظ «إدلر وبنياميني» أن البند كان مؤشراً ممتازاً لمجموعة متنوعة من النتائج الصحية. وقد تفوق أداء هذا البند باستمرار على المتغيرات الأخرى في حساب التباين عبر الدراسات المختلفة. والأكثر ارتباطاً بمناقشتنا، يبدو أنه لا يتعلق بشكل أساسي بالحالة الصحية. وكثيراً ما تتضمن النماذج بنداً واحداً وتضع أيضاً مقاييس للحالة الصحية. وعادةً ما تكون مقاييس التقييم الذاتي ذات البند الواحد للحالة الصحية والمقاييس الأخرى للحالة الصحية منبئين هامين في نفس النموذج. أي إنها لم تشترك في التباين الكافي من أجل المساهمة التنبؤية من جانب واحد لمنع مساهمة تنبؤية مستقلة من الآخر. بل يبدو أن التقييم الذاتي ذا البند الواحد للحالة الصحية يشترك في التباين بدرجة أكبر مع المتغيرات النفسية. وتشير هذه النتائج إلى أن هذا البند الواحد المستخدم على نطاق واسع ليس مؤشراً صادقاً للحالة الصحية، كما يظهر. بالنسبة لهذا البند، فإن النظر إلى ما إذا كان يقيس ما نريد قياسه ليس كافياً لدعم ادعاءات الصدق.

والمشكلة الثانية في تقييم المقياس استناداً إلى الصدق الظاهري هي أن هناك أوقاتاً يكون فيها من المهم ألا يكون المتغير الذي يجري قياسه واضحاً. فعلى سبيل المثال، فإن أداة ما تهدف إلى تقييم الدرجة التي يجيب بها الأشخاص على نحو غير صحيح (كذباً) (على سبيل المثال، لجعل أنفسهم «يبدون جيدين») لن يكون من المفيد أن يكون غرضه واضحاً للمجيبين. هل نستنتج أنه غير صحيح لأنه لا يبدو كما لو كان يقيس عدم الصدق؟ أتمنى ألا يكون كذلك. وهكذا، لدينا هنا حالة لا يمكن فيها أن يؤيد عدم النظر إلى ما هو عليه في الواقع استنتاج البطلان.

ومن الشواغل والاعتبارات الأخيرة فيما يتعلق بالصدق الظاهري أنه ليس من الواضح لمن ينبغي أن يكون الغرض من الأداة واضحاً، في ظاهرها. هل هو المستجيب؟ إذا سأل الطبيب

المريض عما إذا كان هو أو هي أكثر عطشاً من المعتاد، فهل يتوقف صدق هذا السؤال على معرفة المريض بما طُلب منه أو سُئل عنه؟ من الواضح أنه ليس كذلك. هل الشخص الذي يصمم ويبنى الأداة هو الذي ينبغي أن يدرك ويعرف الغرض من الأداة؟ ومن الصعب أن نتصور أن الصلة بين محتوى الأداة والمتغير موضع الاهتمام ليست واضحة بالنسبة لمنشئ أداة ما (ربما باستثناء الحالات التي تتعلق بالصدق التجريبي البحث، والصدق اللانظري atheoretical، والصدق المرتبط بمحك). وإذا تم اعتماد هذا المعنى للصدق الظاهري، فسيتم اعتبار جميع المقاييس صادقة أساساً. وأخيراً، هل هو مجتمع علمي أكبر أو أوسع ينبغي أن يدرك ويعرف غرض أداة ما على أساس مظهرها (ظاهرها)؟ ومن المرجح أن يسفر هذا التفسير عن أدلة متضاربة. قد يبدو بند ما إلى بعض الخبراء كما لو كان يقيس متغيراً واحداً وكأنه يقيس متغيراً آخر من وجهة نظر مجموعة خبراء أخرى مؤهلة بنفس الدرجة. وغالباً ما يبدو أن الأشخاص الذين يدعون أن مقياس ما يتمتع بالصدق أو لا يتمتع بالصدق لأنه يبدو أنه يتمتع بالصدق الظاهري أو لا يتمتع بالصدق الظاهري تستند ادعاءاتهم إلى تصوراتهم الشخصية. أي إنه إذا كان القصد من أداة ما ومظهرها متشابهان لبعضهما البعض، فإنهم يميلون إلى اعتبار أن الأداة تتمتع بالصدق الظاهري؛ خلاف ذلك، فهي ليست كذلك. ويبدو أن ذلك يشكل أساساً ضعيفاً لأي ادعاء بالصدق.

يتم الخلط أحياناً بين الصدق الظاهري وصدق المحتوى نظراً لأن كليهما قد يتعلق بمدى ما يبدو عليه محتوى البند ذي الصلة بالمفهوم البنائي محل الاهتمام. ومع ذلك، هناك اختلاف مهم بينهما يتمثل في أن صدق المحتوى يتم تحديده من ناحية إجراءات محددة، وهذه الإجراءات بشكل عام تكون أكثر تنظيماً ودقة وصرامة مقارنة بالتقييمات غير الرسمية للصدق الظاهري. وكما لوحظ، فقد تتضمن هذه الخطوات جمع الأفكار من المجيبين المحتملين، كما تشمل عادةً الحصول على معلومات من الأدبيات النظرية ذات الصلة ومن الخبراء في المجال قيد البحث. على سبيل المثال، قد يُطلب من خبراء المحتوى تقييم محتوى البند في علاقته بتعريف مفهوم بنائي صريح وواضح. ومن المفترض أن يكون لدى هؤلاء الخبراء إطار مرجعي نظري يدعم أحكامهم. فمجرد ظهور الصلة أو العلاقة لا يكون هو المعيار الوحيد. كما أن الأفراد الذين تُستخدم أحكامهم لتقييم الصدق يكونون واضحين. علاوة على ذلك، فإن الاستنتاج بأن الأداة تتمتع بصدق المحتوى لا يقتصر على تحديد الباحثين أنفسهم. وأخيراً، فإن شفافية المجيب ليست أساساً لتقييم صدق المحتوى

لأنه يتم تقييمه عادةً (على سبيل المثال، من خلال أحكام الخبراء). وبالتالي، فإن الطريقة الرسمية للتحقق من المحتوى يمكن أن تتغلب على قيود ومحددات الاعتماد على الصدق الظاهري.

وتبعاً للظروف والأحوال، قد تكون هناك مزايا أو عيوب في أن يكون الهدف من الأداة واضحاً من مظهره. وكما سنرى في الفصل التالي، غالباً ما تنتج عملية تكوين وتوليد البنود عبارات تشير صراحة إلى المتغير محل الاهتمام، وهذا عادةً ليس أمراً سيئاً. أنا لا أقترح أنه يجب بناء الأدوات بشكل عام بحيث لا يكون القصد أو الهدف منها واضحاً من مظهرها؛ بدلاً من ذلك، فإنني أقترح أنه سواء كان الأمر كذلك أم لا، فليس له علاقة بموضوع الصدق.

تمارين:

1- اذكر مثلاً على كيف أن نفس الارتباط بين مقياس ما وسلوك ما قد يكون مؤشراً أو دلالة على كل من الصدق البنائي أو الصدق المرتبط بمعيار «محك»، وشرح كيف أن (أ) الدوافع وراء حساب الارتباط، و(ب) تفسير هذا الارتباط يختلف تبعاً لنوع الصدق الذي يحاول الباحث تقييمه.

2- افترض أن لدى الباحث مقاييس ورقية وقلم رصاص لقياس اثنين من المفاهيم البنائية هما: احترام أو تقدير الذات والتوافق الاجتماعي. الباحث لديه أيضاً درجات استناداً على مقابلات شخصية على نفس المفهومين البنائيين. كيف يمكن استخدام هذه البيانات في مصفوفة الصفات المتعددة - الأساليب المتعددة لإظهار أن طريقة جمع البيانات كان لها تأثير قوي غير مرغوب فيه على النتائج التي تم الحصول عليها؟

3- كيف يمكن أن يؤثر تعريف نطاق أو مجال المحتوى بشكل واسع للغاية لمقياس ما تأثيراً سلبياً على صدق المحتوى الخاص بهذا المقياس عند استخدامه مع مجتمعات إحصائية معينة؟

## الفصل الخامس

### قواعد إرشادية في تطوير المقياس

حتى الآن، كانت مناقشتنا مجردة إلى حدٍ ما. ونحن الآن ننظر في كيفية تطبيق هذه المعرفة. ويقدم هذا الفصل مجموعة من المبادئ التوجيهية أو الإرشادات المحددة التي يمكن للباحثين استخدامها في وضع وتطوير القياس والمقياس.

#### الخطوة الأولى: حدد بوضوح ما تريد قياسه:

يعتقد العديد من الباحثين أن لديهم فكرة واضحة عما يريدون قياسه فقط ليجدوا أن أفكارهم أكثر غموضاً مما كانوا يعتقدون، وهذا واضح بشكل خادع أو مضلل. وكثيراً ما يحدث هذا الإدراك بعد أن تم بذل جهد كبير في إنشاء وتوليد البنود وجمع البيانات - في الوقت الذي تكون فيه التغييرات أكثر تكلفة بكثير مما لو تم اكتشافه في بداية العملية. هل يجب أن يكون المقياس قائماً على نظرية، أم يجب أن تبدأ في اتجاهات فكرية جديدة؟ لأي درجة من الدقة ينبغي أن يكون عليها المقياس؟ هل ينبغي التأكيد على بعض جوانب الظاهرة أكثر من جوانبها الأخرى؟

#### النظرية كوسيلة مساعدة على تحقيق الوضوح:

كما هو موضح في الفصل الأول، يتطلب التفكير بوضوح في محتوى مقياس ما التفكير جلياً في المفهوم البنائي الذي يتم قياسه. وعلى الرغم من أن هناك العديد من الجوانب التقنية التي ينطوي عليها وضع وتطوير مقياس والتحقق من صدقه، فإنه ينبغي ألا نغفل أهمية أن يكون له أساس سليم في النظريات الأساسية والموضوعية ذات الصلة بالظاهرة المراد قياسها. والغرض من أنواع المقاييس التي هي محور التركيز الرئيسي لهذا الكتاب هو قياس الظواهر المراوغة التي لا يمكن



ملاحظتها مباشرة. ونظرًا لعدم وجود معيار ملموس يمكن على أساسه مقارنة أداء هذا النوع من المقاييس، فمن المهم أن يكون لديك بعض الأفكار الواضحة لتكون بمثابة دليل. ويجب إدراك وتحديد حدود الظاهرة حتى لا ينحرف محتوى المقياس دون قصد إلى مجالات غير مقصودة.

تُعد النظرية مساعداً كبيراً على تحقيق الوضوح. وينبغي دائماً الأخذ في الاعتبار نظريات العلوم الاجتماعية ذات الصلة قبل وضع وتطوير مقياس من النوع الذي تمت مناقشته في هذا الكتاب. وإذا تبين أن النظرية الموجودة لا تقدم أي إرشاد أو توجيه لواضعي أو مطوري المقاييس، فقد يقررون حينئذ أن الاتجاه الفكري الجديد ضرورياً. ومع ذلك، ينبغي أن يكون هذا القرار مستنيراً ومدرّساً، ولا يتم التوصل إليه إلا بعد مراجعة النظرية المناسبة المتعلقة بمشكلة القياس قيد النظر. وحتى لو لم تكن هناك نظرية متوفرة لتوجيه الباحثين، فيجب عليهم أن يضعوا صيغتهم المفاهيمية الخاصة قبل محاولة قياسها إجرائياً. وينبغي على الباحثين، في الأساس، أن يحددوا على الأقل نموذجاً نظرياً أولياً أو مبدئياً يكون بمثابة مرشد أو دليل لوضع وتطوير المقياس. وقد يكون هذا النموذج الأولي بسيطاً مثل مجرد تحديد الصياغة الجيدة والموضوعية للظاهرة التي يسعون إلى قياسها. والأفضل من ذلك إدراج وصف لكيفية ارتباط البناء المفاهيمي الجديد بالظواهر القائمة وقياسها إجرائياً.

#### الخصوصية كوسيلة مساعدة على الوضوح:

وقد يكون مستوى التحديد «الخصوصية» أو العمومية التي يتم بها قياس المفهوم البنائي أمراً مهماً أيضاً. وهناك اتفاق عام في العلوم الاجتماعية على أن المتغيرات سوف ترتبط بقوة مع بعضها البعض عندما تتطابق في مستوى التحديد أو الخصوصية (انظر أجزن وفيشبن Ajzen & Fishbein, 1980، للاطلاع على مناقشة ذلك). وفي بعض الأحيان، يكون الهدف من المقياس أن يرتبط أو يتعلق بسلوكيات أو مفاهيم بنائية محددة للغاية، بينما في أحيان أخرى، يكون المطلوب مقياس أكثر عمومية وشمولية.

كمثال توضيحي للمقاييس التي تختلف في درجة التحديد، نأخذ بعين الاعتبار المفهوم البنائي لموضع التحكم أو الضبط. موقع التحكم هو مفهوم يستخدم على نطاق واسع ويتعلق بتصورات الأفراد حول من أو ما يؤثر على النتائج المهمة في حياتهم. ويمكن تطبيق هذا المفهوم البنائي على نطاق واسع، كوسيلة لشرح أنماط السلوك الشامل أو العالمي التي تغطي العديد من الحالات

والمواقف، أو على نطاق ضيق، للتنبؤ بكيفية استجابة الفرد في سياق محدد للغاية. ويمكن أيضاً وصف مصادر التأثير إما على نطاق واسع أو على نطاق ضيق ومحدد للغاية. وعلى سبيل المثال، فإن مقياس “روتتر” Rotter (1966) لموضع التحكم أو الضبط الداخلي - الخارجي، يهتم بالمقياس على مستوى عام إلى حد ما لهذه التصورات. وثمة بُعد واحد يتراوح بين التحكم الشخصي إلى التحكم من خلال عوامل خارجية يُعد هو الأساس الذي يقوم عليه المقياس، والنتائج التي تركز عليها البنود تكون عامة، مثل النجاح الشخصي. كما أن المصادر الخارجية للتحكم أو السيطرة وُصفت بعبارات عامة أيضاً. العبارات التالية خاصة بموضع التحكم الخارجي من مقياس “روتتر” لموضع التحكم الداخلي - الخارجي: “إن العالم يديره عدد قليل من الأشخاص الموجودين في السلطة، وليس هناك الكثير من الشباب الصغير يمكنهم القيام بذلك”.

وقد طورت ليفنسون (Levenson 1973) مقياس متعدد الأبعاد لموضع التحكم أو الضبط يسمح بوجود ثلاثة مواضع للتحكم هي: الذات، والأشخاص الأقوياء الآخرون، والفرصة أو المصير. وهذا يتيح للباحث أن ينظر إلى مصادر التحكم الخارجية بشكل أكثر تحديداً من خلال وصفها بأنها إما الأشخاص الآخرون الأكثر قوة أو المصير. غير أن النتائج التي ركزت عليها ظلت عامة. مثال على بند من مقياس «ليفنسون» «للمقياس الفرعي الآخرون الأقوياء هو «أشعر أن ما يحدث في حياتي يتم تحديده من قبل الآخرين الأقوياء».

وقد طور والستون، ووالستون، ودي فيليبس Wallston, Wallston, and DeVellis (1978) مقاييس موضع التحكم في الصحة متعددة الأبعاد (MHLC) باستخدام المواضع الثلاثة للتحكم التي اقترحتها «ليفنسون»، مع نتائج خاصة بالصحة، مثل تجنب المرض أو الإصابة بالمرض. وهذا بند كعينة من بنود مقياس الآخرين الأقوياء من مقياس موضع التحكم في الصحة متعددة الأبعاد MFLC هو “أن يكون لدي اتصال منتظم مع طبيبي هو أفضل طريقة لتفادي المرض”. وبعد ذلك طور والستون، وستين، وسميث Smith and ,Stein, Wallston (1994) مقياساً أكثر تحديداً لنتائج قياس موضع التحكم في الصحة سُمي ب (MHLC Form C) والذي يتكون من سلسلة من البنود “نموذج”. ويتيح هذا الإجراء للباحث تحديد أي مشكلة صحية ذات أهمية عن طريق استبدال اسم المرض أو الاضطراب بعبارة “حالي” في كل بند من بنود النموذج. إن أحد البنود كعينة لمقياس الآخرين الأقوياء من

مقياس (MHLC Form C)، حيث يمكن استخدامه في دراسة مرض السكري، هو “إذا رأيت (زرت) طبيبي بانتظام، فأنا أقل عرضة للإصابة بمرض السكري”.

ومن المحتمل أن يكون كل مقياس من مقاييس موضع التحكم التي تكون أكثر تحديداً بصورة تدريجية مفيدة. والأمر الأكثر فائدة يعتمد إلى حد كبير على مستوى النتيجة أو عمومية الموضع المتعلق بالسؤال العلمي الذي يتم طرحه. على سبيل المثال، إذا كان المقصود أو الهدف من مقياس موضع التحكم هو التنبؤ بفئة عامة من السلوك أو ستتم مقارنتها مع متغيرات أخرى لتقييم مفاهيم بنائية على المستوى العام، فقد يكون مقياس “روتر” Rotter هو الخيار الأفضل لأنه هو مقياس عام أيضاً. من ناحية أخرى، إذا كان الباحث مهتماً بالتنبؤ على وجه التحديد بكيفية تأثير المعتقدات حول تأثير أشخاص آخرين على بعض السلوكيات الصحية، فإن مقياس (والستون وآخرون، ١٩٩٤) قد يكون أكثر ملاءمة لأن مستوى التحديد يتطابق مع سؤال البحث. وخلال عملية وضع وتطوير كل من هذه المقاييس، كان لكل من هذه المقاييس إطار مرجعي واضح يحدد مستوى التحديد المناسب، بالنظر إلى الوظيفة المقصودة للمقياس. والنقطة المهمة هي أن مطوري المقياس يجب أن يتخذوا قرار التحديد كقرار نشط وليس مجرد توليد مجموعة من البنود، ومن ثم رؤية الشكل الذي تبدو عليه البنود بعد الحقيقة.

يوضح المثال الخاص بمقياس موضع التحكم الخصوصية أو التحديد فيما يتعلق بالنتائج (على سبيل المثال، كيفية إدارة العالم مقابل مشاكل مرض السكري) وموضع التحكم (أي خارجي بشكل عام مقابل المصير والآخرين الأقوياء بشكل منفصل). ومع ذلك، يمكن أن تختلف خصوصية وتحديد المقياس على طول عدد من الأبعاد، بما في ذلك مجالات أو نطاقات المحتوى (على سبيل المثال، القلق مقابل التكيف النفسي على نطاق أوسع)، أو الإعداد (على سبيل المثال، الاستبيانات المصممة خصيصاً من أجل ملاءمة بيانات عمل معينة)، أو المجتمع الإحصائي (على سبيل المثال، الأطفال مقابل البالغين أو الأفراد العسكريين مقابل طلاب الجامعات).

توضيح ما يجب تضمينه في القياس:

يجب على واضعي ومطوري المقياس أن يسألوا أنفسهم عما إذا كان المفهوم البنائي الذي يرغبون في قياسه مختلفاً ومتميزاً عن غيره من المفاهيم البنائية الأخرى. وكما ذكر آنفاً، يمكن وضع مقاييس بحيث تكون واسعة نسبياً أو ضيقة نسبياً فيما يتعلق بالحالات الذي تنطبق عليها.

وهذا هو الحال أيضاً فيما يتعلق بالمفاهيم البنائية التي تغطيها. يُعد قياس القلق العام أمراً مشروعاً تماماً. وقد يقيس مثل هذا المقياس كلاً من القلق من الاختبار والقلق الاجتماعي على حد سواء. وهذا يُعد أمراً جيداً إذا كان مطابقاً لأهداف واضع (مطور) المقياس أو مستخدم المقياس. ومع ذلك، إذا كان الشخص مهتماً بنوع واحد فقط محدد من القلق، فيجب في هذه الحالة أن يستبعد المقياس جميع الأنواع الأخرى للقلق. يمكن للبنود التي قد «تتقاطع أو تشترك» في مفهوم بنائي ذي صلة (على سبيل المثال، الاستفادة من القلق الاجتماعي عندما يكون موضوع الاهتمام هو قلق الاختبار) أن تكون إشكالية.

في بعض الأحيان، قد تتم الاستفادة من (تأثير) البنود المتشابهة على ما يبدو في قياس مفاهيم بنائية مختلفة تماماً. في مثل هذه الحالات، على الرغم من أن الغرض من المقياس قد يكون قياس ظاهرة واحدة، فإنه قد يكون أيضاً حساساً (مناسباً) لظواهر أخرى. على سبيل المثال، بعض مقاييس الاكتئاب، مثل مقياس الاكتئاب الخاص بمركز الدراسات الوبائية «رادلوف» (Radloff, 1977)، تحتوي على بعض البنود التي تستخدم الجوانب الجسدية للاكتئاب (على سبيل المثال، فيما يتعلق بقدرة المجيب على “الانطلاق والتحرك”). وفي سياق بعض الظروف الصحية، مثل التهاب المفاصل، قد تؤدي هذه البنود إلى أخطاء في جوانب المرض بسبب أعراض الاكتئاب (انظر بلالوك، ديفيليس، براون، ووالستون Blalock, DeVellis, Brown, Wallston & 1989، لمناقشة هذه النقطة تحديداً). قد يختار الباحث الذي يطور مقياساً جديداً للاكتئاب تجنب الجوانب الجسدية إذا كان من المقرر استخدام المقياس مع مجتمعات إحصائية معينة (مثل المرضى بمرض مزمن) أو مع مقاييس أخرى للمفاهيم البنائية الجسدية (مثل قصور الغضروف الدرقي “الوسواس القهري”). وقد يكون من المهم جداً، بطبيعة الحال، أن يُستخدم لأغراض أخرى إدراج وتضمين بنود جسدية، كما هو الحال عندما يتعلق البحث على وجه التحديد بالجوانب الجسدية للتأثير السلبي.

#### الخطوة الثانية: توليد مجموعة من البنود:

بمجرد أن يتم توضيح الغرض من المقياس بصورة محددة، يكون واضع أو مطور المقياس جاهزاً للبدء في بناء وإنشاء أداة القياس بشكل جدي. وتتمثل الخطوة الأولى في إنشاء مجموعة كبيرة من البنود المرشحة للإدراج والتضمين بشكل نهائي في المقياس.

#### اختر البنود التي تعكس الغرض من المقياس:

من الواضح أنه ينبغي اختيار هذه البنود أو إنشاؤها مع مراعاة هدف القياس المحدد. وينبغي أن يوجه هذه العملية الوصف الدقيق لما يهدف المقياس عمله بالضبط. تذكر أن جميع البنود التي تشكل مقياساً متجانساً ينبغي أن تعكس المتغير الكامن الذي تقوم عليه. ويمكن اعتبار كل بند بمثابة اختبار، في حد ذاته، لقوة المتغير الكامن. لذلك، ينبغي أن يعكس محتوى كل بند في المقام الأول المفهوم البنائي محل الاهتمام. وستشكل البنود المتعددة اختباراً أكثر ثباتاً من كل بند من البنود على حدة، ولكن يجب أن يظل كل منها حساساً للدرجة الحقيقية للمتغير الكامن.

من الناحية النظرية، يتم اختيار مجموعة جيدة من البنود بطريقة عشوائية من مجموعة كلية أو شاملة universe of items من البنود المتعلقة بالمفهوم البنائي محل الاهتمام. ويفترض أن تكون المجموعة الشاملة من البنود كبيرة بشكل لا نهائي، وهو ما يحول إلى حد كبير دون أي أمل في التعرف عليه في الواقع واستخراج البنود منها بشكل عشوائي. ومع ذلك، ينبغي أن يوضع هذا الوضع المثالي في الاعتبار. إذا كنت تكتب بنوداً جديدة، كما هو الحال في كثير من الأحيان، يجب أن تفكر بشكل إبداعي في المفهوم البنائي الذي تسعى إلى قياسه. ما هي الطرق الأخرى التي يمكن بها صياغة بند ما للوصول إلى المفهوم البنائي؟ وعلى الرغم من أنه لا ينبغي ألا تتجاوز هذه البنود حدود تعريف المفهوم البنائي المحدد، فإنه ينبغي أن تستنفد هذه الإمكانيات أنواع البنود الواقعة داخل تلك الحدود. ويتم تحديد خصائص المقياس بواسطة البنود التي يتكون منها. وإذا كانت البنود تعكس أو تجسد بشكل ضعيف المفهوم الذي عملت على توضيحه فترة طويلة وشاقة، فإن المقياس لن يعكس بدقة جوهر المفهوم البنائي.

ومن المهم أيضاً أن يكون «الشيء» الذي تشترك فيه البنود مفهوماً بنائياً حقيقياً وليس مجرد فئة. وتذكر مرة أخرى أن نماذجنا لوضع وتطوير المقياس تعتبر البنود مظاهر علنية لمتغير كامن مشترك هو سببها. يتم تحديد الدرجات على البنود المتعلقة بمفهوم بنائي مشترك بالدرجة الحقيقية لهذا المفهوم البنائي. ومع ذلك، كما هو مذكور في الفصل الأول، لمجرد أن البنود تتعلق بفئة category مشتركة، فإن ذلك لا يضمن أن لها نفس المتغير الكامن الأساسي. وغالباً ما تحدد مصطلحات مثل الاتجاهات أو العوائق التي تحول دون الامتثال أو أحداث الحياة فئات المفاهيم البنائية بدلاً من المفاهيم البنائية نفسها. مجموعة من البنود التي ستكون في نهاية المطاف أساس المقياس الأحادي البعد ينبغي ألا تنقسم فقط التركيز على الاتجاهات، على سبيل المثال، بل على اتجاهات محددة، مثل الاتجاهات نحو معاقبة متعاطي المخدرات. يمكن للشخص افتراض تصور

حول سمة من سمات شخص ما - متغير كامن، إذا صح التعبير - من شأنه أن “يسبب” ردود على البنود التي تتناول معاقبة متعاطي المخدرات. ومن التحديات الكبيرة أن نتصور أن إحدى السمات أو الخصائص تفسر الاتجاه بشكل عام. وينطبق الشيء نفسه على الأمثلة الأخرى المذكورة. عادةً ما تكون عوائق الامتثال من أنواع عديدة. قد يمثل كل نوع (على سبيل المثال، الخوف من اكتشاف الأعراض، والقلق بشأن تكاليف العلاج، وتوقع الألم، والمسافة إلى مرافق (مؤسسات) العلاج، وتصورات المناعة) متغير كامن. قد يكون هناك حتى علاقات متبادلة بين بعض المتغيرات الكامنة. ومع ذلك، فإن كلاً من هذه العوائق هي مفهوم بنائي منفصل. ومن ثم، فإن مصطلح العوائق يصف فئة من المفاهيم البنائية بدلاً من مفهوم بنائي على حدة يرتبط بمتغير كامن واحد. ولا ينبغي أن يُتوقع من البنود التي تقيس مختلف المفاهيم البنائية التي تندرج ضمن نفس الفئة (مثل التصورات حول المناعة والمخاوف بشأن تكاليف العلاج) أن تختلف في طريقة عمل البنود عندما تكون مظهراً لمتغير كامن مشترك.

#### التكرار:

من المفارقات أو التناقضات أن التكرار Redundancy يُعد سمة أو خاصية جيدة وسيئة للبنود الموجودة في المقياس. ويستلزم حل هذه المفارقة التمييز بين سمات أو خصائص البنود التي تعزز وتدعم المقياس من خلال التكرار وتلك التي لا تفعل ذلك. لأن هذا الموضوع غالباً ما يكون مصدرًا للالتباس، فسوف أناقش ذلك بشيء من التفصيل، وسأقوم بادئ ذي بدء بالقضية التي تؤدي التكرار.

في هذه المرحلة من عملية وضع وتطوير المقياس، من الأفضل أن يكون أكثر شمولاً، وأن تكون كل الأمور الأخرى متساوية. التكرار ليس بالأمر السيء عند وضع المقياس. في الواقع، فإن النماذج النظرية التي توجه جهودنا لوضع وتطوير المقاييس تعتمد على التكرار. في مناقشة معادلة تنبؤ “سبيرمان براون Spearman-Brown” في الفصل الثالث، أشرت إلى أن الثبات يختلف كدالة في عدد البنود، وكل الأمور الأخرى تكون متساوية. ونحن نحاول تجسيد الظاهرة محل الاهتمام من خلال وضع وتطوير مجموعة من البنود التي تكشف عن هذه الظاهرة بطرق مختلفة. وباستخدام بنود متعددة تبدو متكررة، فإن المحتوى المشترك بين البنود سيتم جمعه عبر البنود بينما سيتم إلغاء الخواص غير ذات الصلة. وبدون التكرار، سيكون ذلك مستحيلاً.

ومع ذلك، فإن كل أشكال التكرار أمراً ليس مرغوباً فيه. ويتعلق التكرار المفيد بالمفهوم البنائي، وليس بالجوانب العرضية أو الثانوية للبند. تأمل في البندين التاليين: النسخة الأصلية هي «شيء مهم حقاً نجاح طفلي»، والنسخة المعدلة، «الشيء المهم حقاً هو نجاح طفلي». لم يتم تغيير شيء أكثر من أداة النكرة «a» إلى أداة التعريف «the» في بند ما بالتأكيد سيعطيك تكراراً فيما يتعلق بالمحتوى الهام للبند، ولكن البنود الأصلية والمعدلة ستكون أيضاً تكراراً فيما يتعلق بالعديد من الأشياء التي تريد تغييرها، مثل بنيتها النحوية الأساسية واختيار الكلمات. ومن ناحية أخرى، فإن بندين مثل «سأفعل أي شيء تقريباً لضمان نجاح طفلي» و«لا توجد تضحية أكبر إذا كان ذلك يساعد طفلي على النجاح» قد تكون تكراراً بشكل مفيد لأنها تعبر عن فكرة مماثلة بطرق مختلفة إلى حد ما. فهي تكرار فيما يتعلق بالمتغير محل الاهتمام ولكنها ليست تكراراً فيما يتعلق ببنيتها النحوية ومفرداتها العرضية (الثانوية). وعندما يتم تجنب التكرارات التي ليست ذات صلة، فإن التكرارات ذات الصلة سوف تسفر عن مجموعات بنود أكثر ثباتاً.

علاوة على ذلك، وعلى الرغم من أن التكرار في الأداة النهائية قد يكون غير مرغوب فيه، فإن المسألة تكون أقل أهمية خلال المراحل المبكرة من وضع البنود. وبناءً على ذلك، قد تكون هناك قيمة وفائدة لصيغتي أو نسختي البند المختلفتين في كلمة واحدة فقط لإدراجهما وتضمينهما في اختبار البند الأولي. من خلال عمل ذلك، يمكن التحقق مما إذا كانت إحدى النسختين للبند متفوقة (أي أكثر جودة) على الأخرى، ومن ثم يمكن إدراج وتضمين البند المتفوق في النسخة النهائية للمقياس. وقد تم تقديم حجة ضد التكرار جزئياً وهي: أن سمات البنود السطحية المكررة مثل المفردات العرضية أو الثانوية (أي غير ذات الصلة بالمفهوم البنائي) أو البنية النحوية لا تُعد ميزة. قد تؤدي أوجه التشابه في الصياغة التي لا صلة لها بالمفهوم البنائي إلى تفاعل المجيبين بشكل مماثل مع البنود بطريقة تسفر عن تقدير مبالغ فيه للثبات. على سبيل المثال، إذا كانت هناك عدة بنود تبدأ بعبارة مشتركة (على سبيل المثال، «عندما أفكر في الأمر، ...»)، قد يؤدي مجرد التشارك في تلك العبارة إلى ارتباط تلك البنود ببعضها البعض بشكل أقوى. ولا يميز مؤشر الثبات مثل «ألفاكرونباخ» بين التباين للبنود الناشئ عن تلك الصياغة المشتركة والتباين الذي يعزى إلى التأثير المشترك للمتغير محل الاهتمام. ومن ثم، سيكون تقدير الثبات مبالغاً فيه أو متضخماً.

وفي حين أن السمات النحوية المماثلة أو غيرها من السمات السطحية الأخرى يمكن أن تشكل تشابهاً غير مرغوب فيه في المحتوى، قد يؤدي التكرار غير المرتبط بشكل كامل بالمفهوم

البنائي موضع الاهتمام إلى حدوث مشكلة في بعض الظروف. ويمكن أن يحدث هذا عندما تختلف بعض البنود عن معظم البنود الأخرى في مجموعة فيما يتعلق بالخصوصية أو التحديد. على سبيل المثال، لنفكر ملياً في مقياس افتراضي يهدف إلى قياس الاتجاهات نحو محبي الحيوانات الأليفة. قد يكون من المناسب إدراج مجموعة متنوعة من البنود في المقياس. وقد تكون هناك بنود أخرى أيضاً، وإن كانت ذات صلة بالمفهوم البنائي موضع الاهتمام، محددة أكثر من اللازم - وبالتالي مكررة إلى حد كبير - للعمل بشكل جيد. قد تكون البنود التالية «عشاق البيغاء الرمادي الأفريقي هم أشخاص لطاف» و«أعتقد أن الناس الذين يحبون البيغاوات الرمادية الأفريقية هم أشخاص طيبون» متشابهة إلى حد كبير ليس فقط بسبب تشابه البنية النحوية ولكن بسبب احتمال أن يكون المحتوى الذي يتقاسمه البندين ملائماً ولكنه محدد بشكل مبالغ فيه. قد يقومون بسحب البند الذي تم تحديده ككل بعيداً عن المتغير الكامن المستهدف أو المقصود (الاتجاهات نحو محبي الحيوانات الأليفة) إلى بديل أكثر تحديداً للمتغير الكامن (الاتجاهات نحو عشاق البيغاء الرمادي الأفريقي). ومع الأخذ في الاعتبار وجود مجموعة كبيرة من الحيوانات الأليفة، هناك بندان حول نوع معين وغير مألوف من الحيوانات الأليفة متشابهان بشكل كبير، ومن المحتمل أن يقوضا الغرض من الأداة.

وبصورة أعم، فإن مدى عمومية أو تحديد المفهوم البنائي موضع الاهتمام يمكن أن يغير من تأثير التكرار. وعلى الرغم من أن مثال البيغاء الرمادي الأفريقي قد يبدو متطرفاً بعض الشيء، فإن إدراج بنود لا تتطابق مع خصوصية أو تحديد المفهوم البنائي موضع الاهتمام يمكن أن يحدث في سياق أقل غرابة. على سبيل المثال، في أداة تم تصميمها لتعكس أو تجسد جميع جوانب العاطفة، فإن العديد من البنود المتعلقة بالقلق قد تمثل مشكلة. ومن المحتمل أن تكون الارتباطات بين هذه البنود أكبر من الارتباطات بين تلك البنود وغيرها من البنود التي لا تتعلق بالقلق. وكنتيجة لذلك، قد تشكل هذه البنود مجموعة فرعية من بنود القلق داخل المجموعة الأوسع من بنود العاطفة. وهذا يمكن أن يسبب عدداً من المشكلات. أولاً، قد يؤدي ذلك إلى تقويض أو الإضرار بأحادية البعد لمجموعة البنود (التي ستكون مشكلة إذا كانت نية أو هدف الباحث هو وضع مقياس واحد لمتغير أحادي البعد). أيضاً، قد يؤدي ذلك إلى إنشاء نقطة اتصال غير مقصودة ينتج عنها بنود أكثر تشابهاً لتلك الموجودة في عنقود (مجموعة) القلق التي يبدو أنها أفضل أداءً من تلك البنود التي تكون أقل تشابهاً. فعلى سبيل المثال، على الرغم من أن بنداً متعلقاً بالقلق وآخر متعلق بالخوف قد يكونان على نفس القدر من الأهمية بالنسبة لوجهة النظر الواسعة للعواطف، فإن البند الأول قد يسهم بقوة



أكبر في الثبات من البند الأخير إذا كانت هناك أغلبية عددية لبنود القلق في الأداة. ونتيجة لذلك، قد يتجاوز متوسط ارتباط بند القلق متوسط ارتباط بند الخوف، مما يؤدي إلى زيادة مساهمته في تقدير الثبات. والواقع أن الأداة التي تتضمن تمثيلاً مبالغاً فيه لبنود القلق على النحو الموصوف لن تعبر عن جميع العواطف بالتساوي ولكن ستكون منحازة نحو القلق.

وعلى النقيض من ذلك، فإن نفس النوع من بنود القلق التي توصف بأنها تمثل إشكالية في الفقرة السابقة قد لا تكون تكراراً بشكل مفرط أو مبالغ فيه في أداة ذات تركيز أضيق. ومن الواضح أنه إذا كانت الأداة مصممة لتقييم القلق، فينبغي أن تكون جميع البنود ذات صلة بذلك المتغير وأن التشابه لن يكون حالة من حالات التكرار غير المرغوب فيه. وعلى النقيض من ذلك، فإن البنود التي تتضمن عبارة أكثر عمومية، مثل «مشاعري العامة»، قد تشكل مجموعة فرعية إذا تم تضمينها في حالة مقياس القلق بسبب تركيزها العاطفي غير المحدد. غير أن ما يبدو على أنه مشكلة تكرار قد يكون في الواقع مسألة مدى تطابق البنود مع خصوصية وتحديد المفهوم البنائي الذي يعتزم الباحث تقييمه.

في أداة تهدف إلى تجسيد متغير أكثر تحديداً، فمن المرجح أن البنود سوف تظهر أو تبدو أكثر تشابهاً مع بعضها البعض. عادةً، على سبيل المثال، ستبدو البنود الموجودة في مقياس يقيس القلق في مجال الخطابة العامة أكثر تشابهاً مع بعضها البعض (نظراً لخصوصية المتغير موضع الاهتمام) من البنود الموجودة في مقياس يقيس الحالات العاطفية على نطاق أوسع. وقد لا تكون هذه مشكلة طالما كانت أوجه التشابه مرتبطة بالمفهوم البنائي موضع الاهتمام. وكما ذكرنا من قبل، فإن البنود المتشابهة بقدر ارتباطها بمتغيرها المقصود وليس في أي مجال آخر يمكن أن تكون بنوداً جيدة.

#### عدد البنود:

من المستحيل تحديد عدد البنود التي ينبغي إدراجها أو تضمينها في تجمع أولي. ويكفي القول إنك تريد أكثر بكثير مما تخطط لإدراجه في المقياس النهائي. تذكر أن ثبات الاتساق الداخلي هو دالة لمدى قوة ارتباط البنود مع بعضها البعض (وبالتالي مع المتغير الكامن) وعدد البنود الموجودة لديك في المقياس. وبما أن طبيعة الارتباطات المتبادلة بين البنود تكون غير معروفة عادةً في هذه المرحلة من وضع وتطوير المقياس، فإن وجود الكثير من البنود يُعد شكلاً من أشكال

التأمين ضد ضعف الاتساق الداخلي. كلما زاد عدد البنود التي لديك في المجموعة المتوفرة لديك في تجمع البنود، كان الأمر أكثر صعوبة بشأن اختيار البنود التي ستؤدي المهمة التي تنوي القيام بها. وليس من غير المعتاد البدء بمجموعة من البنود التي يبلغ حجمها ثلاثة أو أربعة أضعاف الحجم النهائي للمقياس. وهكذا، قد يتطور مقياس مكون من عشرة بنود من مجموعة تضم 40 بنوداً. وإذا كان من الصعب بشكل خاص توليد وإنشاء بنود في مجال معين من مجالات المحتوى أو إذا كانت البيانات التجريبية تشير إلى أنه لا توجد حاجة إلى بنود عديدة لتحقيق اتساق داخلي جيد، فقد يكون التجمع الأولي للبنود أصغر بنسبة 50% من حجم المقياس النهائي.

بشكل عام، كلما كان تجمع البنود أكبر، كان ذلك أفضل. ومع ذلك، فمن المؤكد أنه من الممكن تطوير تجمع كبير جداً من البنود بحيث لا يمكن إدارته في مناسبة واحدة لأي مجموعة واحدة من الأشخاص. إذا كان تجمع البنود كبيراً بشكل استثنائي، يمكن للباحث أن يلغي ويحذف بعض البنود وفقاً لمعايير مسبقة، مثل عدم الوضوح، والأهمية أو الصلة المشكوك فيها، أو التشابه غير المرغوب فيه مع بنود أخرى.

#### بدء عملية كتابة البنود:

غالبًا ما يكون البدء في كتابة البنود هو أصعب جزء في عملية توليد وتكوين البنود. واسمحوا لي أن أصف كيف أبدأ هذه العملية. في هذه المرحلة، أكون أقل اهتماماً بجودة البنود ونوعيتها من مجرد التعبير عن الأفكار ذات الصلة. وأبدأ في كثير من الأحيان بعبارة أو فقرة تمثل إعادة صياغة للمفهوم البنائي الذي أريد قياسه. على سبيل المثال، إذا كنت مهتماً بوضع وتطوير مقياس للإدراك الذاتي لتأثير الإعلانات التجارية، فيمكن أن أبدأ بالعبارة أو الفقرة التالية «أنا سريع التأثير بالإعلانات التجارية». بعد ذلك سأحاول توليد وإنشاء فقرات وعبارات إضافية تدور حول نفس الفكرة بشكل مختلف إلى حد ما. قد تكون العبارة التالية «الإعلانات التجارية تؤثر عليّ كثيراً». وسأستمر بهذه الطريقة أو الأسلوب، مع عدم الاهتمام بوجود أي معايير للجودة تقريباً في العبارات. وهدف في هذه المرحلة المبكرة هو مجرد تحديد مجموعة كبيرة من الطرق التي يمكن من خلالها توضيح وتحديد المفهوم الأساسي الذي تستهدفه الأداة. وأثناء الكتابة، فإنني قد أبحث عن طرق بديلة للتعبير عن الأفكار النقدية. على سبيل المثال، قد أستبدل عبارة «الإعلانات التجارية» بعبارة هي «الأشياء التي أراها في إعلانات التلفزيون أو المجلات» في المجموعة التالية من

الجمال. وأنا أجد أن الكتابة بسرعة ودون نقد يُعد أمراً مفيداً. بعد توليد وإنشاء ربما ثلاثة أو أربعة أضعاف عدد البنود التي أتوقع إدراجها وتضمينها في الأداة النهائية، سوف ألقى نظرة على ما قمت بكتابته. والآن قد حان الوقت الحاسم أو المهم. حيث يمكن فحص البنود للوقوف على مدى جودة تجسيدها للأفكار الأساسية ومدى وضوح الصياغة. وتحدد الأقسام التالية بعض خصائص البنود المحددة لتجنبها أو إدراجها في عملية الاختيار من قائمة العبارات الأصلية وتنقيحها.

#### خصائص البنود الجيدة والبنود السيئة:

إن سرد جميع الأشياء والأمور التي تجعل بنداً ما جيداً أو سيئاً تعد مهمة مستحيلة. ومن الواضح أن مجال أو نطاق المحتوى له تأثير كبير محتمل على جودة البند. ومع ذلك، هناك بعض الخصائص التي تفرق وتميز على نحو موثوق أفضل البنود عن أسوأ البنود. وتتعلق معظم هذه الخصائص بالوضوح. وكما هو مبين في الفصل الأول، ينبغي أن يكون البند الجيد واضحاً لا لبس فيه. وينبغي التخلي عن الأسئلة التي تضع المجيب في مأزق وارتباك.

يجب على واضعي أو مطوري المقياس تجنب البنود الطويلة بشكل استثنائي، حيث إن طول البند عادةً ما يزيد من التعقيد ويقلل من الوضوح. غير أنه ليس من المرغوب فيه أو المستصوب التوضيح بمعنى أو بمدلول بند ما لصالح الإيجاز. إذا كان تعديل البند ضرورياً لإيصال ونقل الغرض من أحد البنود، فقم بإدراجه. ومع ذلك، تجنب الإسهاب والإطناب غير الضروري. وبوجه عام، سيكون بند ما مثل «كثيراً ما أجد صعوبة في التعبير عن وجهة نظري» أفضل من عبارة أطول دون داع، مثل «من الإنصاف القول إن أحد الأشياء التي يبدو لي أنني أواجه فيها مشكلة في كثير من الوقت هو إيصال ونقل وجهة نظري إلى أشخاص آخرين».

وثمة اعتبار آخر ذي صلة باختيار أو تطوير البنود هو مستوى صعوبة القراءة الذي تتم به كتابة البنود. هناك مجموعة متنوعة من الأساليب (على سبيل المثال، دايل وشل، Fry, 1977؛ 1948 Chall, & Dale) لتعيين مستويات الصف بالنسبة للكلام النثري بما في ذلك بنود المقياس. عادةً ما تساوي هذه الكلمات والجمال الأطول مع مستويات القراءة الأعلى. ويفترض أن قراءة معظم الصحف المحلية تتطلب مستوى قراءة للصف السادس.

يحدد "فري" (Fry 1977) عدة خطوات لتحديد مستوى القراءة كميًا. الخطوة الأولى هي تحديد عينة من النص الذي يبدأ بالكلمة الأولى من الجملة ويحتوي على 100 كلمة بالضبط. (بالنسبة للمقاييس التي تحتوي على عدد قليل من البنود فقط، قد تضطر إلى تحديد جزء مناسب من 100 وتأسيس الخطوات اللاحقة على هذه النسبة.) بعد ذلك، احسب عدد الجمل الكاملة والمقاطع اللفظية الفردية في عينة النص. ويتم استخدام هذه القيم كنقاط إدخال لمخطط أو رسم بياني يوضح مستوى الصف المكافئ أو المعادل لتوليفات وتركيبات مختلفة من الجملة وتعداد المقاطع اللفظية من العينة المكونة من 100 كلمة. يشير الرسم البياني إلى أن متوسط عدد الكلمات والمقاطع لكل جملة لمستوى القراءة في الصف الخامس هو 14 و18 على التوالي. وتبلغ الجملة المتوسطة في الصف السادس 15 أو 16 كلمة وما مجموعه 20 مقطعاً لفظياً؛ أما جملة مستوى الصف السابع فتتكون من حوالي 18 كلمة و24 مقطعاً لفظياً. ويمكن أن تسفر الجمل الأقصر التي تحتوي على نسبة أعلى من الكلمات الأطول أو الجمل الأطول مع عدد أقل من الكلمات الطويلة عن مستوى يعادل الصف. على سبيل المثال، يتم تصنيف جملة مكونة من 9 كلمات و13 مقطعاً لفظياً (أي ما يصل إلى 44% كلمات متعددة المقاطع أو متشابكة) أو جملة مكونة من 19 كلمة و22 مقطعاً لفظياً (أي، ما لا يزيد عن 14% من الكلمات متعددة المقاطع أو متشابكة) على حد سواء على أنها تعادل مستوى القراءة للصف السادس. ومن المحتمل أن يكون الهدف من تحقيق مستوى للقراءة بين الصفين الخامس والسابع هدفاً مناسباً لمعظم الأدوات التي سيتم استخدامها مع المجتمع الإحصائي العام. فعلى سبيل المثال، كُتبت بنود مقاييس موضع الضبط أو السيطرة الصحية المتعدد الأبعاد على مستوى القراءة من الصف الخامس إلى الصف السابع. وهناك بند نموذجي في هذا المستوى من القراءة هو «معظم الأشياء التي تؤثر على صحتي تحدث لي عن طريق الصدفة» (Wallston et al., 1978). هذا البند يتكون من 11 كلمة و15 مقطعاً لفظياً مما يضع هذا البند في مستوى القراءة للصف السادس.

ويشير فري (Fry 1977) إلى أنه ينبغي النظر في العوامل الدلالية والنحوية وأخذها في الاعتبار عند تقييم صعوبة القراءة. ونظرًا لأن الكلمات القصيرة تميل إلى أن تكون أكثر شيوعًا، وتميل الجمل القصيرة إلى أن تكون أبسط من الناحية النحوية، فإن الإجراء الخاص بـ "فري" يعد بديلاً مقبولاً لطرق تقييم الصعوبة الأكثر تعقيداً. ومع ذلك، وكما هو الحال مع المعايير الأخرى للكتابة أو اختيار البنود الجيدة، يجب على الشخص استخدام المنطق السليم في تطبيق أساليب مستوى القراءة. بعض العبارات المختصرة التي تحتوي فقط على كلمات قصيرة ليست أولية أو

بسيطة. على سبيل المثال، عبارة مثل «تجنب الازدراء في الكاسك» Eschew casque scorn من المرجح أن تربك شخصاً ما بمستوى تعليم ابتدائي أكثر من عبارة “ارتدِ خوذتك” Wear your helmet، على الرغم من أن كل جملة يوجد بها ثلاث كلمات وأربعة مقاطع لفظية. وهناك مصدر آخر للارتباك والالتباس المحتمل يجب تجنبه وهو كلمات النفي «السلبيات» المتعددة. العبارة التالية «أنا لا أؤيد قيام الشركات بوقف تمويل الجماعات المناهضة للأسلحة النووية» تُعد أمراً مربكاً أكثر بكثير من العبارة التالية «أنا أؤيد الدعم الخاص المستمر للمجموعات التي تدعو إلى حظر الأسلحة النووية». (من المفيد أيضاً ملاحظة أن هاتين العبارتين قد تعبران عن مواقف مختلفة بشأن هذه القضية. فعلى سبيل المثال، قد تنطوي هذه العبارة الأخيرة على تفضيل دعم القطاع الخاص عن دعم القطاع العام للجماعات المعنية).

وينبغي أيضاً تجنب ما يسمى بالبند المزدوج double-barreled items. وهي البنود التي تنقل فكرتين أو أكثر بحيث أن الإجابة على البند بالموافقة قد يشير إلى أي من الأفكار أو كليهما. العبارة التالية «أنا أؤيد الحقوق المدنية لأن التمييز وعدم المساواة جريمة ضد الله» هي مثال على البند المزدوج. إذا كان الشخص يدعم الحقوق المدنية لأسباب أخرى غير الإساءة إلى الذات الإلهية (على سبيل المثال، لأنها جريمة ضد الإنسانية)، فكيف ينبغي أن يجيب/ تجيب على البند؟ فقد تعبر الإجابة السلبية بشكل غير صحيح عن عدم تأييد الحقوق المدنية، والإجابة الإيجابية قد تنسب بشكل غير صحيح دافع إلى دعم المجيب.

وثمة مشكلة أخرى يجب على واضعي ومطوري المقياس تجنبها هي إشارات الضمير الشخصي pronoun الغامضة. البند التالي “يجب على القتل ومرتكبي جرائم الاغتصاب ألا يلتمسوا العفو من السياسيين لأنهم حثالة الأرض” قد يعبر عن مشاعر بعض الناس بغض النظر عن إشارة الضمير. (ومع ذلك، عادةً ما ينوي مطور مقياس أن يكون أكثر وضوحاً حول ما يعنيه البند). هذه الجملة يجب أن تكون سيئة أو غير مرغوبة مرتين. فبالإضافة إلى إشارة الضمير الغامضة، فهي بند مزدوج. وتؤدي المُعدِّلات modifiers التي تكون في غير محلها إلى حالات غموض والتباس مماثلة لإشارات الضمير الغامضة: الجملة التالية: “ينبغي لأعضائنا في الكونغرس أن يعملوا بجد لإضفاء الشرعية (لتقنين البغاء) على البغاء في مجلس النواب” هي مثال على هذه المُعدِّلات. استخدام أشكال أو نماذج الصفات بدلاً من أشكال الأسماء يمكن أيضاً أن يحدث

ارتباكاً وتشويشاً غير مقصود. النظر في الاختلافات في المعنى بين All vagrants should be given a schizophrenic assessment “ينبغي إعطاء جميع المتشردين تقييم المصاب بانفصام الشخصية” و All vagrants should be given a schizophrenia assessment “ينبغي إعطاء جميع المتشردين تقييم انفصام الشخصية”.

إن الكلمات الفردية ليست هي المصادر الوحيدة لغموض البند. فالجملة بأكملها يمكن أن يكون لها أكثر من معنى واحد. ولم أرَ في الواقع سوى دراسة استقصائية واحدة للسلوك الجنسي للمراهقين تضمنت بنداً لتقييم تعليم الوالدين. وبالنظر إلى سياق الدراسة الاستقصائية ككل، فإن الصياغة كانت مؤسفة: «إلى أي مدى ذهبت أمك إلى المدرسة؟» (How far did your mother go in school?) وقد فشل الباحثون تماماً في التعرف على المعنى غير المقصود لهذه الجملة إلى الحد الذي أثار سخط مجموعة من المهنيين أثناء عرض حلقة النقاش. وأظن أن عدداً لا بأس به من المجيبين المراهقين قد ضحكوا أيضاً من هذا البند. ولا يُعرف كيف أثر ذلك على ردودهم على ما تبقى من الاستبانة.

#### البند ذات الصياغة الإيجابية والبند ذات الصياغة السلبية:

يختار العديد من مطوري المقاييس كتابة البنود ذات الصياغة السلبية التي تمثل مستويات منخفضة أو حتى غياب المفهوم البنائي موضع الاهتمام وكذلك البنود ذات الصياغة الإيجابية الأكثر شيوعاً، والتي تمثل وجود المفهوم البنائي موضع الاهتمام. والهدف من ذلك هو الوصول إلى مجموعة من البنود، يشير بعضها إلى مستوى عالٍ من المتغير الكامن عند الموافقة عليه وإقراره، وأخرى تشير إلى مستوى عالٍ عند عدم الموافقة على المتغير الكامن. على سبيل المثال، يتضمن مقياس «روزنبرغ» (Rosenberg (1965 لتقدير الذات، على سبيل المثال، بنوداً تشير إلى التقدير العالي للذات (على سبيل المثال، “أشعر بأن لديّ عدداً من الصفات الجيدة”) وأخرى تشير إلى تقدير للذات منخفض (على سبيل المثال، “أشعر بالتأكد أنني عديم الفائدة في بعض الأحيان”). والغرض من صياغة البنود بشكل إيجابي أو سلبي على حد سواء ضمن نفس المقياس هو عادةً تجنب القبول أو التأكيد أو التحيز في الموافقة. تشير هذه المصطلحات التي تستخدم بالتبادل إلى ميل المجيب إلى الموافقة على البنود بغض النظر عن محتواها ومضمونها. إذا كان المقياس، على سبيل المثال، يتكون من بنود تعبر عن درجة عالية من تقدير الذات، فإن تحيز القبول قد يؤدي إلى نمط

من الاستجابات التي تبدو وكأنها تشير إلى تقدير للذات مرتفع للغاية. وإذا كان المقياس مكوناً من عدد متساوٍ من البنود ذات الصياغة الإيجابية والسلبية، فمن ناحية أخرى، يمكن التمييز بين تحيز الموافقة والدرجة المتطرفة من تقدير الذات عن بعضهما البعض عن طريق نمط الاستجابات. من شأن “الموافقة” أن يؤيد البنود التي تشير إلى تقدير الذات العالية والمنخفضة على حد سواء، في حين أن الشخص الذي كان يحظى بتقدير كبير حقاً سيؤيد بشدة البنود التي تحظى بتقدير كبير ويؤيد بصورة سلبية البنود ذات التقدير المنخفض.

ولسوء الحظ، قد يكون هناك ثمن يجب دفعه مقابل إدراج وتضمين البنود ذات الصياغة الإيجابية والسلبية. قد يكون تغيير الاتجاه (تحويل الصياغة السالبة إلى موجبة والعكس) في قطبية العناصر مربكة للمستجيبين، خاصة عند ملء استبيان طويل. وفي مثل هذه الحالة، قد يصبح المستجيبون في حيرة من أمرهم حول الفرق بين التعبير عن قوة اتفاقهم مع جملة (بند) ما، بغض النظر عن قطبيتها، وبين التعبير عن قوة السمة التي يتم قياسها (التقدير، على سبيل المثال). وبصفتي باحثاً في العلوم الاجتماعية التطبيقية، فقد رأيت العديد من الأمثلة للبنود التي صيغت في الاتجاه المعاكس تؤدي إلى أداء سيء للبنود. على سبيل المثال، وصف «ديفيليس وكالاهان» (DeVellis and Callahan 1993) بديلاً أقصر وأكثر تركيزاً لمؤشر الاتجاهات نحو مرض الروماتيزم (وهو اسم مؤسف، لأن الأداة لا تقيّم الاتجاه ولا تُعد مؤشراً). لقد اخترنا بنوداً من النسخة الأصلية والأطول بناءً على معايير تجريبية وانتهى بنا الأمر إلى أربعة بنود تعبر عن ردود الفعل السلبية على المرض وأحدهم يعبر عن القدرة على مواجهة (التغلب على) المرض بشكل جيد. وكان القصد من ذلك هو أن يقوم المستخدمون بعكس درجة تسجيل بند “المواجهة” بحيث تكون جميع البنود معبرة عن الشعور بالعجز. وفي الأونة الأخيرة، درس «كوري، كالاهان، وديفيليس» (DeVellis and Callahan, Currey 2002) أداء هذا العنصر الفردي الذي صيغ في الاتجاه الإيجابي. وقد كان أدائه ضعيفاً (منخفضاً) باستمرار. وعندما تمت إعادة صياغة البند ببساطة عن طريق إضافة كلمة (لا) لتغيير تكافؤه بحيث يكون متسقاً مع البنود الأخرى، فقد تحسن أدائه بشكل كبير. ونعتقد أنه على الرغم من إدراك العديد من المجيبين للتكافؤ المختلف للبند الأصلي، إلا أن آخرين لم يدركوا ذلك. ومن شأن ذلك أن يؤدي إلى وجود نسبة من الأفراد الذين أكدوا الارتباطات الإيجابية مع البنود الأربعة الأخرى ونسبة أخرى لهم نفس الارتباطات السلبية. ونتيجة لذلك، بالنسبة للعينة ككل، فإن الارتباطات بين هذا البند والبنود الأربعة الأخرى سوف تتضاءل أو

تتخفّض بشكل ملحوظ، وبالتالي، سوف ينتج نوع الأداء غير المرضي الذي لاحظناه بالنسبة للبند الأصلي عكس بند التكافؤ. توحى لي التجربة أو الخبرة الشخصية مع العينات المعتمدة على المجتمع المحلي أن عيوب ومساوئ البنود التي صيغت في اتجاه معاكس تفوق أي فوائد.

#### الخلاصة:

يجب أن يكون تجمع pool البنود مصدراً غنياً يمكن أن ينشأ منه مقياس. وينبغي أن يتضمن عدداً كبيراً من البنود ذات الصلة بالمحتوى موضع الاهتمام. التكرار فيما يتعلق بالمحتوى هو أصل، وليس التزام. وهو أساس ثبات الاتساق الداخلي الذي هو، بدوره، أساس الصدق. وينبغي ألا تنطوي البنود على “صفقة شاملة أو مجموعة مقاييس” تجعل من المستحيل على المجيبين تأييد جزء من البند دون تأييد جزء آخر قد لا يكون متسقاً مع الجزء الأول. سواء تم تضمين البنود ذات الصياغة الإيجابية والسلبية في تجمع البنود أم لا، فإن صياغتها يجب أن تتبع القواعد اللغوية المعمول بها. وسوف يساعد ذلك على تجنب بعض مصادر الغموض التي نوقشت أعلاه.

#### الخطوة الثالثة: تحديد شكل أو تنسيق القياس:

توجد أشكال وصيغ عديدة للأسئلة. يجب على الباحث أن يفكر مبكراً في الشكل الذي سيكون عليه المقياس. وينبغي أن تحدث هذه الخطوة بشكل متزامن مع عملية توليد وإنشاء البنود بحيث يتم التوافق بين الاثنين. على سبيل المثال، قد يكون إنشاء وتكوين قائمة طويلة من العبارات التوضيحية مضيعة للوقت إذا كان شكل الاستجابة الذي وقع عليه الاختيار في نهاية المطاف عبارة عن قائمة مرجعية مؤلفة من بنود مكونة من كلمة واحدة. علاوة على ذلك، النماذج النظرية المقدمة في وقت سابق هي أكثر اتساقاً مع بعض صيغ أو أشكال الاستجابة من غيرها. وبصفة عامة، فإن المقاييس التي تتكون من بنود قابلة للتسجيل على مقياس متصل ويتم تجميعها لتشكل درجة المقياس هي الأكثر توافقاً مع التوجه النظري الوارد في هذا الكتاب. ومع ذلك، سأناقش في هذا القسم الأشكال والصيغ الشائعة التي تختلف عن النمط الذي تتضمنه النماذج النظرية التي تمت مناقشتها في الفصل الثاني وكذلك الأشكال التي تنقيد بهذا النمط.

#### مقياس ثرستون (Thurstone):



هناك عدد من الإستراتيجيات العامة لبناء المقاييس تؤثر على شكل البنود وخيارات الاستجابة. إحدى هذه الإستراتيجيات هي مقياس «ثورستون»<sup>Thurstone</sup>. قد يساعد التشبيه على توضيح كيفية عمل مقياس «ثورستون». تم تصميم شوكة رنانة لتهتز بتردد معين. إذا قمت بضربها، فإنها ستهتز في هذا التردد وستؤدي إلى إصدار نغمة معينة. وعلى العكس من ذلك، إذا وضعت الشوكة بالقرب من مصدر نغمة ينتج نفس تردد شوكة الرنين، تبدأ الشوكة بالاهتزاز. بمعنى أن الشوكة الرنانة عبارة عن «جهاز كشف تردد»، تهتز في وجود موجات صوتية من التردد الرنان وتظل بدون حركة في وجود كل الترددات الأخرى. تخيل سلسلة من الشوكات الرنانة المرتبة في مصفوفة ما بحيث عندما يتحرك إحداها من اليسار إلى اليمين على طول المصفوفة، تتوافق الشوكات الرنانة مع أصوات التردد الأعلى تدريجيًا. ضمن نطاق تردد الشوكة الرنانة، يمكن استخدام هذا الترتيب لتحديد تردد النغمة. وبعبارة أخرى، يمكنك تحديد تردد النغمة من خلال رؤية أي شوكة تهتز عند تشغيل النغمة. إن مقياس «ثورستون» تم إعداده ليعمل بنفس الطريقة. يحاول واضع ومطور المقياس توليد وإنشاء بنود تستجيب بشكل متفاوت لمستويات معينة من السمة المعنية. عندما تتطابق «درجة» بند معين مع مستوى السمة التي يمتلكها المجيب، فإن هذا البند سيشير إلى هذا التطابق. وفي كثير من الأحيان، تتألف الإشارة من استجابة إيجابية للبنود التي يتم «ضبطها» على المستوى المناسب من السمة واستجابة سلبية لجميع البنود الأخرى. وعادةً ما يتم تحديد عملية الضبط (أي تحديد مستوى المفهوم البنائي الذي يستجيب له كل بند) عن طريق قيام المحكمين بوضع مجموعة كبيرة من البنود في مجموعات تقابل بفواصل أو فترات زمنية متساوية من حيث أهمية المفهوم البنائي أو قوته.

هذه فكرة ممتازة للغاية. ويمكن تطوير البنود لتتوافق مع درجات شدة مختلفة من السمة، ويمكن أن تكون متباعدة بحيث تمثل فواصل أو فترات متساوية، ويمكن تنسيقها وصياغتها باستخدام خيارات الاستجابة موافق وغير موافق، على سبيل المثال. ويمكن للباحث إعطاء هذه البنود للمجيبين ثم يفحص إجاباتهم لمعرفة البنود التي أدت إلى الاتفاق. ونظراً لأن البنود ستتم معايرتها مسبقاً فيما يتعلق بحساسيتها لمستويات محددة من الظاهرة، فإن الاتفاقات ستحدد مقدار السمة التي يمتلكها المجيب. قد يؤدي اختيار البنود التي تمثل فترات متساوية عبر البنود إلى خصائص قياس مرغوب فيها للغاية لأن الدرجات ستكون قابلة للإجراءات الرياضية القائمة على القياس الفكري أو المسافي.

قد يبدو جزء من مقياس «ثورستون» الافتراضي لقياس تطلعات الآباء للتحصيل التعليمي والوظيفي لأطفالهم كما يلي:

- تحقيق النجاح هو السبيل الوحيد لطفلي أن يسدد الجهود التي بذلتها كوالد.	موافق ....	غير موافق ....
- الذهاب إلى كلية جيدة والحصول على عمل جيد أمر مهم ولكن ليس ضرورياً لسعادة طفلي.	موافق ....	غير موافق ....
- السعادة ليست لها علاقة بتحقيق الأهداف التعليمية أو المادية.	موافق ....	غير موافق ....
- زخارف النجاح ذات القيمة المعتادة ليست عائقاً أمام السعادة الحقيقية.	موافق ....	غير موافق ....

وكما يشير نونالي (1978) Nunnally، فإن تطوير مقياس ثورستون حقيقي وصحيح أصعب بكثير من وصفه. وأنه من الصعب للغاية العثور على البنود التي «يتردد صداها» باستمرار على مستويات محددة من هذه الظاهرة. وغالباً ما تفوق المشاكل العملية المرتبطة بهذه الطريقة مزاياها ما لم يكن لدى الباحث سبب مقنع لرغبته في نوع المعايرة التي يوفرها. على الرغم من أن مقياس «ثورستون» هو طريقة أو نهج مثير للاهتمام ومناسب في بعض الأحيان، إلا أنه لن تتم الإشارة إليه في بقية هذا الكتاب. ومع ذلك، لاحظ أن الأساليب القائمة على نظرية استجابة البند، التي تمت مناقشتها في فصل لاحق، تشترك في العديد من أهداف مقياس «ثورستون» مع اتباع نهج مختلف إلى حد ما لتحقيقها.

مقياس جتمان (Guttman):

مقياس «جتمان» Guttman هو سلسلة من البنود التي تؤثر تدريجياً على مستويات أعلى من السمة. وهكذا، ينبغي للمجيب أن يؤيد ويوافق على مجموعة من البنود المتجاورة إلى أن يتجاوز، عند نقطة حرجية، مقدار السمة التي تستخدمها البنود ويمتلكها الشخص. ولا يجب تأييد وإقرار أي من البنود المتبقية. وتتوافق بعض البيانات الوصفية البحتة مع مقياس «جتمان». على

سبيل المثال، قد تطرح سلسلة من أسئلة المقابلة، “هل تدخن؟” “هل تدخن أكثر من 10 سجائر في اليوم؟” “هل تدخن أكثر من علبة سجائر في اليوم؟” وهكذا. كما هو الحال في هذا المثال، فإن إقرار أي بند محدد على مقياس “جتمان” يعني ضمناً تأكيد جميع البنود السابقة. ويشير أعلى بند يعطي رداً إيجابياً إلى مستوى السمة الذي يمتلكه المجيب. لاحظ أنه، في حين أن كلاً من مقياس «ثرستون» Thurstone ومقياس “جتمان” Guttman يتكونان من بنود متدرجة، فإن التركيز ينصب على استجابة إيجابية واحدة في الحالة الأولى، ولكن في الحالة الأخيرة ينصب التركيز على نقطة الانتقال أو التحول من الاستجابات الإيجابية إلى السلبية في الحالة الأخيرة. قد تبدو نسخة مقياس “جتمان” من مقياس الطموح الأبوي السابق كما يلي:

- تحقيق النجاح هو السبيل الوحيد لطفلي لسداد جهودي كوالد.	موافق ....	غير موافق ....
- الذهاب إلى كلية جيدة والحصول على عمل جيد أمرٌ مهمٌ ولكن ليس ضرورياً لسعادة طفلي.	موافق ....	غير موافق ....
- السعادة ليست لها علاقة بتحقيق الأهداف التعليمية أو المادية.	موافق ....	غير موافق ....
- زخارف النجاح ذات القيمة المعتادة ليست عائناً أمام السعادة الحقيقية.	موافق ....	غير موافق ....

يمكن أن تعمل مقاييس “جتمان” بشكل جيد للحصول على معلومات موضوعية أو في الحالات التي يكون فيها الضرورة المنطقية هي أن الاستجابة الإيجابية لمستوى واحد من التسلسل الهرمي تعني استيفاء معايير جميع المستويات الأدنى من هذا التسلسل الهرمي. وتصبح الأمور أكثر غموضاً عندما لا تكون الظاهرة محل الاهتمام ملموسة. وفي حالة مقياسنا الافتراضي عن الطموح الأبوي، على سبيل المثال، قد لا يكون الترتيب موحداً بين الأفراد. في حين أن تدخين 20 سيجارة في اليوم تعني دائماً تدخين أكثر من 10 سجائر، والردود على البندين رقم (٣) ورقم (٤) في مثال مقياس الطموح الأبوي قد لا تتوافق دائماً مع نمط الترتيب في مقياس «جتمان». فعلى سبيل المثال، قد يوافق الشخص على البند رقم (٣) ولكنه لا يوافق على البند رقم (٤). ومن الطبيعي، أن تنطوي الموافقة على البند رقم (٣) الموافقة على البند رقم (٤)، ولكن إذا رأى المجيب

أن النجاح عامل معقد أو مركب يعمل في آن واحد كمساعد وعائق أمام السعادة، فإن ذلك قد يسفر عن نمط غير عادي من الاستجابات.

مثل مقياس «ثورستون»، فإن مقياس «جتمان» بلا شك له مكانته، ولكن إمكانية تطبيقهما تبدو محدودة إلى حد ما. ومع كلا النهجين أو النوعين من المقاييس، غالباً ما تفوق المساوئ والمصاعب المزايا. ومن المهم أيضاً التأكيد من جديد على أن نظريات القياس التي تمت مناقشتها حتى الآن لا تنطبق دائماً على هذه الأنواع من المقاييس. ومن المؤكد أن افتراض وجود علاقات سببية قوية بنفس القدر بين المتغير الكامن وكل بند من البنود لن ينطبق على مقياس «ثورستون» أو مقياس «جتمان». يصف «نونالي وبرنشتاين» (Nunnally and Bernstein (1994) بإيجاز بعضاً من النماذج المفاهيمية التي تستند إليها هذه المقاييس. أما بالنسبة للحالات التي تكون فيها البنود المرتبة ملائمة بشكل خاص، فإن النماذج المستندة إلى نظرية استجابة البند (التي تمت مناقشتها في الفصل السابع) قد تكون خياراً مناسباً، على الرغم من أن تنفيذ هذه الأساليب يمكن أن يكون مرهقاً للغاية.

#### المقاييس ذات البنود متساوية الترجيح أو الوزن:

تتناسب نماذج القياس التي تمت مناقشتها سابقاً بشكل أفضل مع المقاييس التي تتكون من بنود تعادل إلى حد ما «أجهزة كشف» للظاهرة محل الاهتمام - أي إنها متكافئة بدرجة أو بأخرى (ولكنها ليست متكافئة بالضرورة بالمعنى الدقيق لنموذج الاختبارات المتوازية «المتكافئة»). وهي مؤشرات غير كاملة لظاهرة مشتركة يمكن دمجها عن طريق الجمع البسيط في مقياس يتمتع بالثبات إلى حد مقبول.

وهناك ميزة واحدة جذابة للمقاييس من هذا النوع هي أن كل بند من البنود على حدة يمكن أن يكون له مجموعة متنوعة من أشكال وتنسيقات خيارات الاستجابة، وهذا يتيح لمطور المقياس قدراً كبيراً من المدى أو الطول في إنشاء مقياس مناسب بشكل مثالي للغرض المحدد. وسيتم أدناه بحث بعض القضايا والمشكلات العامة المتعلقة بشكل وصيغة الاستجابة، وكذلك بحث مزايا وخصائص بعض أشكال وصيغ الاستجابة النموذجية.

#### كم عدد فئات الاستجابة؟

تتكون معظم بنود المقياس من جزأين هما: جذع أو ساق، وسلسلة من خيارات الاستجابة (السلم). على سبيل المثال، قد يكون الساق لكل بند فقرة توضيحية وتفسيرية مختلفة يعبر عن رأي ما، وقد تكون خيارات الاستجابة المصاحبة لكل ساق عبارة عن سلسلة من الأوصاف التي تشير إلى قوة الاتفاق مع البند. دعونا نركز الآن على خيارات الاستجابة - وعلى وجه التحديد، عدد الخيارات التي ينبغي أن تكون متاحة للمجيب. تسمح بعض أشكال وصيغ استجابة البنود للشخص بوجود عدد لا نهائي أو كبير جداً من الخيارات، في حين أن بعضها الآخر يحد من الاستجابات المحتملة. تخيل، على سبيل المثال، تدرج الاستجابة لمقياس الغضب الذي يشبه مقياس الحرارة، تمت معايرته من «لا غضب على الإطلاق» في قاعدة المقياس إلى «الغضب التام الذي لا يمكن السيطرة عليه» في قمته. ويمكن تقديم مجموعة من توصيفات الحالة أو الموقف إلى المجيب، مصحوبة كل منها بنسخة من المقياس، ويطلب منه أن يبين، عن طريق التظليل في جزء من المقياس، مدى أو مقدار الغضب الذي تثيره الحالة أو الموقف. تسمح هذه الطريقة بالمقياس المتصل «المستمر» للغضب فعلياً. وقد تطلب طريقة بديلة من المجيب أن يبين، باستخدام رقم يتراوح بين درجة واحدة و100 درجة، مدى ومقدار الغضب الذي يسببه له كل موقف أو حالة. وهذا يوفر العديد من الاستجابات المنفصلة. وبدلاً من ذلك، قد يؤدي الشكل أو التنسيق إلى تقييد خيارات الاستجابة إلى عدد قليل من الخيارات، مثل «لا شيء» أو «قليلاً» أو «مقدار معتدل أو متوسط» و«كثيراً» أو إلى اختيار ثنائي بسيط بين «غاضب» و«غير غاضب».

ما هي المزايا النسبية لهذه البدائل؟ أحد خصائص النوعية أو الجودة المرغوب فيها لتدرج القياس هو الاختلاف والتباين. فالمقياس لا يمكن أن يكون متبايناً إذا لم يختلف أو يتغير. وإذا فشل مقياس ما في التمييز بين الاختلافات في السمة الأساسية، فإن ارتباطاته بالمقاييس الأخرى ستكون مقيدة وسوف تكون فائدته محدودة. ومن بين الطرق لزيادة فرص التباين هي إدراج وتضمين المقياس عدداً كبيراً من البنود. وثمة خيار آخر يتمثل في توفير خيارات عديدة للاستجابة في إطار البنود. إذا اضطرت الظروف الباحث إلى أن يقتصر على سؤالين فقط يتعلقان بقياس الغضب، على سبيل المثال، فقد يكون من الأفضل السماح للمجيبين بمزيد من حرية التعبير في وصف مستوى غضبهم. افترض أن البحث يتعلق بتطبيق سياسات عدم التدخين في بيئة العمل. ولنفترض كذلك أن الباحثين يريدون تحديد العلاقة بين هذه السياسة والغضب. إذا كانت تقتصر على سؤالين فقط (على سبيل المثال، «ما مقدار الغضب الذي تشعر به عندما يتم حظرك من التدخين؟» و«ما مقدار

الغضب الذي تشعر به عندما تتعرض للآخرين ممن يدخلون في مكان العمل؟»، وقد يحصل الباحثون على معلومات أكثر فائدة من صيغ الاستجابة التي تسمح للأشخاص بتدريج استجابة أكثر من صيغ الاستجابة الثنائية (على سبيل المثال، «غاضب» و«ليس غاضب»). فعلى سبيل المثال، قد يكشف مقياس من «صفر» إلى «100» عن اختلافات كبيرة في ردود الفعل إزاء هذه الحالات والمواقف كما قد يؤدي إلى تباين جيد في مقياس مكون من بندين. ومن ناحية أخرى، إذا سُمح لفريق البحث بإدراج «50» سؤالاً عن التدخين والغضب، فإن المؤشرات البسيطة «الغاضبة» مقابل «غير الغاضبة» قد تسفر عن تباين كافٍ عندما تضاف أو يتم تجميع البنود للحصول على درجة المقياس. وفي الواقع، فإن مواجهة المزيد من خيارات الاستجابة بشأن كل سؤال من الأسئلة الخمسين قد تؤدي إلى إرهاق المجيبين أو تتسبب في إزعاجهم ومللهم، مما يقلل من ثبات إجاباتهم.

وثمة مسألة أخرى تتعلق بعدد خيارات الاستجابة تتمثل في قدرة المجيبين على التمييز بشكل كبير. ما مدى دقة التمييز الذي يمكن أن يقوم به الشخص المثالي؟ ومن الواضح أن هذا يتوقف على ما يتم قياسه. وهناك عدد قليل من الأشياء يمكن تقييمها حقاً في 50 فئة منفصلة، على سبيل المثال. مع عرض هذه الخيارات العديدة، قد لا يمكن للعديد من المجيبين استخدام سوى تلك الخيارات المقابلة لمضاعفات العدد خمسة أو عشرة، مما يقلل بشكل فعال عدد الخيارات إلى عدد قليل لا يتجاوز خمسة. وقد لا تعكس الاختلافات بين استجابة من 35 إلى 37 الفرق أو الاختلاف الفعلي في الظاهرة التي يتم قياسها. ولا يكسب هذا النوع من الدقة الزائفة إلا القليل. وعلى الرغم من أن تباين المقياس قد يزداد، فإنه قد يكون الجزء العشوائي الذي يتزايد (أي الخطأ) وليس الجزء المنتظم الذي يعزى إلى الظاهرة الأساسية. وهذا، بطبيعة الحال، لا يقدم أو يحقق أي فائدة.

وفي بعض الأحيان، تعتمد قدرة المجيب على التمييز بشكل مفيد بين خيارات الاستجابة على الصياغة المحددة لتلك الخيارات أو موضعها (مكانها) المادي. وقد تؤدي مطالبة المجيب بالتمييز بين التوصيفات الكمية الغامضة، مثل «عدة» و«قليل» و«كثير»، إلى خلق مشاكل. في بعض الأحيان، يمكن تقليل الغموض عن طريق ترتيب خيارات الاستجابة على الصفحة. ويبدو أن المجيبين غالباً ما يفهمون ما هو مرغوب فيه عندما تُعرض عليهم سلسلة متصلة واضحة. وبالتالي، فإن الطلب مثل:

كثيراً	بعض الأحيان	قليلاً	قليلاً جداً	أبداً
--------	-------------	--------	-------------	-------

وقد يعني هذا ضمناً أن الخيار «بعض الأحيان» هو أكثر من خيار الاستجابة «قليلاً» بسبب ترتيب هذه البنود. ومع ذلك، إذا كان من الممكن إيجاد صفة لا لبس فيها تمنع المجيبين من وضع افتراضات تستند إلى الموضع أو المكان على طول سلسلة متصلة، فإن هذا يكون أفضل بكثير. وفي بعض الأحيان، قد يكون من الأفضل أن يكون لديك عدد من خيارات الاستجابة أقل من تلك التي تكون غامضة. لذلك، على سبيل المثال، قد يكون من الأفضل في المثال السابق حذف إما «بعض الأحيان» أو «قليلاً» ويكون لديك أربعة خيارات بدلاً من خمسة. وأسوأ الظروف هي الجمع بين الكلمات الغامضة ومواقع الصفحة الغامضة. فكر في المثال التالي:

مفيد جداً Very Helpful	غير مفيد جداً Not Very Helpful
مفيد إلى حد ما Somewhat Helpful	غير مفيد على الإطلاق Not at All Helpful

ومن الصعب جداً التمييز بين مصطلحات مثل إلى حد ما somewhat و ليس إلى حد كبير not very في ظل أفضل الظروف. ومع ذلك، فإن ترتيب خيارات الاستجابة هذه كما تظهر أعلاه يجعل الأمور أكثر سوءاً. إذا قام المجيب بقراءة العمود الأول ثم أسفل العمود الثاني، فيبدو أن تعبير «إلى حد ما somewhat» يمثل قيمة أعلى من تعبير «ليس إلى حد كبير not very». ولكن إذا قرأ المجيب الصف الأول ثم عبر الصف الثاني، فإن الترتيب الضمني لهذين الاستجابتين الواصفتين على طول سلسلة متصلة يتم عكسه. ونظرًا للغموض الذي يكتنف كلاً من الترتيب اللغوي والترتيب المكاني، يمكن للأفراد أن يعطوا معاني مختلفة للخيارين اللذين يمثلان قيمتين معتدلتين، وكنتيجة لذلك، فإن الثبات سيتأثر.

وثمة مسألة أخرى تتمثل في قدرة الباحث واستعداده لتسجيل عدد كبير من القيم لكل بند. إذا تم استخدام طريقة مقياس الحرارة التي تم وصفها سابقاً لتحديد الاستجابات الكمية على مقياس الغضب، فهل سيحاول الباحث بالفعل تسجيل درجات دقيقة لكل استجابة؟ ما هي الدقة المناسبة؟ هل يمكن قياس المنطقة المظلمة في حدود ربع بوصة؟ سنتيمتر؟ مليمتراً؟ إذا تم استخراج بعض البيانات الخام فقط - مثل الثلث الأدنى أو الأوسط أو الأعلى - من المقياس، فما هي النقطة التي كانت عندها المطالبة بمثل هذه الاستجابة الدقيقة؟

وهناك مسألة أو مشكلة أخرى على الأقل تتعلق بعدد الردود أو الاستجابات. على افتراض السماح بعدد قليل من الإجابات المنفصلة لكل بند، هل ينبغي أن يكون عدد الردود أو الاستجابات فردياً أم زوجياً؟ ومرةً أخرى، يعتمد هذا على نوع السؤال، ونوع خيار الاستجابة، وغرض أو هدف الباحث. وإذا كانت خيارات الاستجابة ثنائية القطب، حيث يشير أحد القطبين إلى عكس القطب الآخر (على سبيل المثال، اتجاه إيجابي قوي مقابل اتجاه سلبي قوي)، فإن عدداً فردياً من الردود والاستجابات يسمح (بالمراوغة) بالغموض واللبس (على سبيل المثال، “لا أتفق ولا أعارض neither agree nor disagree”) أو عدم اليقين أو عدم التأكد (على سبيل المثال، “غير متأكد”)؛ وعادةً لا يسمح العدد الزوجي للاستجابات بذلك. وينطوي العدد الفردي للاستجابات على نقطة “محايدة” مركزية (على سبيل المثال، لا تقييم إيجابي ولا سلبي). من ناحية أخرى، هناك عدد زوجي من الردود أو الاستجابات يجبر المجيب على أن يقدم على الأقل التزاماً ضعيفاً في اتجاه أحد الطرفين أو الطرف الآخر (مثل الاختيار القسري بين التقييم الإيجابي إلى حد ما أو التقييم السلبي إلى حد ما باعتباره الرد الأقل تطرفاً). ولا يكون أي من الشكلين متفوقاً بالضرورة. وقد يرغب الباحث في منع (المراوغة) الغموض واللبس إذا كان هناك شعور بأن الأشخاص سيختارون استجابة محايدة كوسيلة لتجنب الاختيار. ففي الدراسات المتعلقة بخيارات المقارنة الاجتماعية، على سبيل المثال، قد يرغب الباحثون في إجبار الأشخاص على التعبير عن تفضيلهم للحصول على معلومات حول الشخص الذي يتمتع بقدر أكبر من المزايا أو بقدر أقل من المزايا. لننظر في هذين الشكلين البديلين، الأول الذي تم اختياره لدراسة المقارنات الاجتماعية بين المصابين بالتهاب المفاصل (DeVellis et al., 1990):

١- هل تفضل معلومات حول:

(أ) المرضى الذين يعانون من التهاب المفاصل أسوأ مما لديك.

(ب) المرضى الذين يعانون من التهاب المفاصل أكثر اعتدالاً مما لديك.

٢- هل تفضل معلومات حول:

(أ) المرضى الذين يعانون من التهاب المفاصل أسوأ مما لديك.

(ب) المرضى الذين يعانون من التهاب المفاصل بنفس القدر الذي تعاني منه.



(ج) المرضى الذين يعانون من التهاب المفاصل أكثر اعتدالاً مما لديك.

قد يسمح خيار محايد مثل رقم ٢ (ب) بالمرادغة (اللبس والغموض) غير المرغوب فيها. وقد يكون من المطلوب والمستحسن أيضاً وجود نقطة محايدة. في دراسة تقييم أي نوع من المخاطر (على سبيل المثال، الملل مقابل الخطر) الذي يفضلها الناس، قد تكون نقطة الوسط ضرورية أو حاسمة. قد يغير الباحث فرصة أو شدة الضرر عبر عدة خيارات بين نشاط آمن وممل ونشاط مثير ومحفوف بالمخاطر. ويمكن عندئذ استخدام النقطة التي يكون فيها المجيب ملتبساً إلى أبعد حد بشأن المخاطرة بالنشاط الأكثر إثارة كمؤشر للمخاطرة:

يرجى الإشارة إلى تفضيلك النسبي للنشاط (أ) أو النشاط (ب) من بين البدائل المدرجة أدناه من خلال وضع دائرة حول العبارة المناسبة بعد وصف النشاط (ب).

النشاط (أ): قراءة كتاب إحصائي (لا توجد فرصة لحدوث إصابة شديدة).

١- النشاط (ب): القيام برحلة جوية في طائرة ركاب صغيرة (احتمال ضئيل جداً بحدوث إصابة بالغة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)
-------------------------	-------------------------------	---------------------	-------------------------------	-------------------------

٢- النشاط (ب): القيام برحلة على متن طائرة صغيرة في قمرة القيادة (فرصة طفيفة للإصابة الشديدة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)
-------------------------	-------------------------------	---------------------	-------------------------------	-------------------------

٣- النشاط (ب): القفز بالمظلة من طائرة مع مظلة احتياطية (فرصة بشكل معتدل لإصابة شديدة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)
-------------------------	-------------------------------	---------------------	-------------------------------	-------------------------

٤- النشاط (ب): القفز بالمظلة من طائرة بدون مظلة احتياطية (خطر كبير من الإصابة الشديدة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)
-------------------------	-------------------------------	---------------------	-------------------------------	-------------------------

٥- النشاط (ب): القفز من طائرة بدون مظلة ومحاولة الهبوط على هدف ناعم (إصابة خطيرة مؤكدة تقريباً).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)
-------------------------	-------------------------------	---------------------	-------------------------------	-------------------------

وبصرف النظر عن المزايا أو الالتزامات الأخرى لهذا النهج أو الأسلوب، فإنه يتطلب بوضوح أن تتضمن خيارات الاستجابة نقطة وسط أو نقطة منتصف.

أنواع محددة (معينة) من أشكال وتنسيقات الاستجابة:

تظهر بنود المقياس في مجموعة متنوعة من الأشكال والصيغ المثيرة. ومع ذلك، هناك عدة طرق لعرض وتقديم البنود التي يتم استخدامها على نطاق واسع وأثبتت نجاحها في تطبيقات متنوعة. وناقش أدناه بعض هذه المسائل.

مقياس ليكرت:

يُعد مقياس «ليكرت» Likert scale أحد الأشكال والصيغ الأكثر شيوعاً للبنود. وعند استخدام مقياس «ليكرت»، يتم عرض البند كجملة تفسيرية، تليها خيارات استجابة تشير إلى

درجات مختلفة من الاتفاق مع العبارة أو تأييدها. (في الواقع، كان المثال السابق عن المخاطرة يستخدم شكل وصيغة استجابة مقياس ليكرت). وتبعاً للظاهرة التي يجري دراستها وأهداف الباحث، فقد يصاحب كل بند إما عدد فردي أو عدد زوجي من خيارات الاستجابة. وينبغي صياغة خيارات الاستجابة بحيث تكون لها فترات متساوية تقريباً فيما يتعلق بالاتفاق. وهذا يعني أن الفرق في الاتفاق بين أي زوج متجاوز من الإجابات يجب أن يكون مماثلاً تقريباً لأي زوج متجاوز آخر من خيارات الاستجابة. ومن الممارسات الشائعة إدراج ستة ردود أو استجابات محتملة هي: «غير موافق بشدة»، و«غير موافق إلى حد ما»، و«غير موافق بشكل معتدل أو متوسط»، و«موافق بشكل معتدل أو متوسط»، و«موافق إلى حد ما»، و«موافق بشكل معتدل أو متوسط». وهذا يشكل سلسلة متصلة من عدم الاتفاق بشدة إلى الاتفاق بشدة. يمكن أيضاً إضافة نقطة وسط أو منتصف محايدة. وتشمل الخيارات الشائعة لنقطة الوسط «لا أوافق ولا أعارض» و«موافق وغير موافق بالتساوي (سيان)». فهناك مجال مشروع للمناقشة فيما يتعلق بتكافؤ هاتين النقطتين المتوسطتين. الأول يعني ضمناً عدم الاكتراث واللامبالاة، في حين يشير الأخير إلى وجود جاذبية قوية ولكنها متساوية لكل من الاتفاق وعدم الاتفاق. وقد يكون الأمر جيداً أن معظم المحبيين لا يركزون كثيراً على الفروق الدقيقة في اللغة، بل يعتبرون مجرد خيار الاستجابة المعقول في وسط النطاق كنقطة وسط أو منتصف، بصرف النظر عن صياغته الدقيقة.

يستخدم مقياس «ليكرت» على نطاق واسع في أدوات قياس الآراء والمعتقدات والاتجاهات. ومن المفيد غالباً أن تكون هذه العبارات قوية إلى حد ما (وإن لم تكن قوية للغاية) عند استخدامها في شكل وصيغة مقياس ليكرت. ومن المفترض أن يتم التعبير عن الاعتدال في الرأي في اختيار خيار الاستجابة. على سبيل المثال، العبارات التالية: «يتجاهل الأطباء عمومًا ما يقوله المرضى»، و«في بعض الأحيان، لا يولي الأطباء القدر من الاهتمام الذي ينبغي لتعليقات المرضى»، و«مرة واحدة من حين لآخر، قد ينسى الأطباء أو يفوتهم شيء ما أخبرهم به المريض» هي تعبر عن آراء قوية وآراء معتدلة وآراء ضعيفة، على التوالي، فيما يتعلق بعدم اهتمام الأطباء بتعليقات المرضى. أيهم يكون أفضل لمقياس «ليكرت»؟ في نهاية المطاف، بطبيعة الحال، فإن الذي يعكس الاختلافات الحقيقية في الرأي بشكل أكثر دقة يكون هو الأفضل. عند اختيار مدى قوة الصياغة في تجمع أولي للبنود، قد يكون من المفيد أن يتساءل الباحث عن «الطريقة التي من المحتمل أن يجيب بها الناس الذين لديهم مقادير أو نقاط قوة مختلفة في السمة المعنية» وفي حالة الأمثلة الثلاثة المقدمة للتو، فإن

الباحث قد يستنتج أن السؤال الأخير ربما قد يؤدي إلى التوصل إلى اتفاق قوي على الأرجح من جانب الأشخاص الذين تتدرج آراؤهم على امتداد جزء كبير من السلسلة المتصلة من الإيجابية إلى السلبية. وإذا ثبت أن هذا الاستنتاج صحيح، فلن تؤدي العبارة الثالثة دوراً جيداً في التمييز بين الأشخاص ذوي الآراء السلبية القوية مقابل الآراء السلبية المعتدلة (المتوسطة).

وبصفة عامة، قد تؤدي العبارات المعتدلة للغاية إلى التوصل إلى اتفاق أكثر مما ينبغي عند استخدامها في مقاييس «ليكرت». وسوف يتفق كثير من الناس بشدة مع عبارة مثل «سلامة وأمن المواطنين أمر مهم». ويمكن للشخص أن يوافق بشدة على مثل هذه العبارة (أي اختيار خيار الاستجابة القصوى) دون إبداء رأي مبالغ فيه. وبطبيعة الحال، فإن العكس هو الصحيح بنفس القدر. وقد يجد الأشخاص الذين لديهم أي آراء أكثر تطرفاً أنفسهم في خلاف (عدم اتفاق) مع عبارة قوية للغاية (على سبيل المثال، «إن ملاحقة مرتكبي الأفعال غير المشروعة ومعاقتهم أهم من حماية حقوق الأفراد»). ومن بين العبارتين (الأكثر اعتدالاً والأكثر تطرفاً)، فإن الأولى قد تكون هي المشكلة الأكبر لسببين. أولاً، غالباً ما نميل إلى كتابة عبارات لا تسيء إلى الناس. إن تجنب الهجوم قد يكون فكرة جيدة. ومع ذلك، فإنه قد يقودنا إلى تفضيل البنود التي سيجدها الجميع تقريباً مقبولة. وهناك سبب آخر يدعو إلى الحذر من البنود المعتدلة جداً، وهو أنها قد تمثل غياب أو عدم وجود الاعتقاد أو الرأي. أما البند الثالث من بنودنا الخاصة بالطبيب الذي لا ينتبه لتعليقات المرضى في الفقرة السابقة فلم يبين وجود اتجاه إيجابي بقدر ما يدل على عدم وجود اتجاه غير إيجابي. وقد لا تتناسب بنود من هذا النوع بدرجة كافية مع هدف البحث لأننا غالباً ما نهتم بوجود ظاهرة ما أكثر مما نهتم بعدم وجودها.

تتمثل إحدى الطرق المفيدة لمعايرة مدى قوة أو اعتدال صياغة عبارة ما في القيام بما يلي: تخيل المستجيب العادي الذي يكون هدفاً لتطبيق المقياس. حاول أن تتخيل كيف سيستجيب هذا الشخص على بنود مختلفة من حيث القوة. والآن، فكر في نوع صياغة البند الذي من المرجح أن يستخلص إجابة من المجيب العادي الذي كان في مركز خيارات الاستجابة لمقياس «ليكرت» والذي تخطط لاستخدامه أو بالقرب منه. لذلك، على سبيل المثال، إذا اخترت مقياساً ذا تدرج مكون من 6 نقاط وكان فيه خيار الاستجابة المركزي (الوسط) هو «غير موافق قليلاً وموافق قليلاً»، فإنك تريد تكوين بند يستخلص إحدى هذه الاستجابات من مجيب عادي في المجتمع الإحصائي محل الاهتمام. وينبغي أن يكون مثل هذا البند قادراً على استيعاب الأشخاص الذين تكون وجهات نظرهم (آراؤهم)

إما أقل أو أكثر قوة من المجيب العادي أو المتوسط، وبالتالي ينبغي من الناحية النظرية أن يكون لديهم تباين كبير في الدرجة الملاحظة والدرجة الحقيقية. وبفضل التباين بين المجيبين، ستتاح للبند فرصة أفضل للارتباط بشكل جيد مع البنود الأخرى (لأن التباين المشترك «التغاير» مرتبط بمدى التباين) وبالتالي سيكون لديه القدرة على تعزيز ثبات المقياس. وعلى النقيض من ذلك، فإن البند الذي من المحتمل أن تنتج عنه استجابات شديدة «متطرفة أو مبالغ فيها»، مثل عدم الموافقة بشدة أو الموافقة بشدة، من المجيب العادي، سوف يؤدي عملاً ضعيفاً في التمييز بين مختلف الفئات الكاملة من المجيبين.

باختصار، يجب أن يوضح بند مقياس «ليكرت» الجيد الرأي أو الاتجاه أو المعتقد أو أي مفهوم بنائي آخر قيد الدراسة بعبارات واضحة. وليس من الضروري ولا من المناسب أن يمتد هذا النوع من المقياس إلى نطاق التأكيدات الضعيفة إلى تأكيدات قوية للمفهوم البنائي. توفر خيارات الاستجابة الفرصة للتدرجات.

فيما يلي أمثلة للبنود بصيغ وأشكال استجابة مقياس «ليكرت»:

١- تعتبر ممارسة التمارين الرياضية عنصراً أساسياً في نمط حياة صحي.

١	٢	٣	٤	٥	٦
غير موافق بشدة	غير موافق إلى حد ما	غير موافق بشكل معتدل	موافق بشكل معتدل	موافق إلى حد ما	موافق بشدة

٢- ينبغي أن تكون مكافحة تعاطي المخدرات أولوية وطنية عليا أو قصوى.

١	٢	٣	٤	٥
صحيح تماماً	صحيح في الغالب	صحيح و غير صحيح على حد سواء	غير صحيح في الغالب	غير صحيح تماماً

## القياس التفاضلي الدلالي (تباين الدلالات اللفظية):

ترتبط طريقة القياس التفاضلي الدلالي بشكل رئيسي ببحوث الاتجاه التي أجراها «أوسغود وزملاؤه» (على سبيل المثال. Osgood & Tannenbaum, 1955). وعادةً، ما يستخدم تباين الدلالات اللفظية في الإشارة إلى واحد أو أكثر من الدوافع أو المحفزات. ففي حالة الاتجاهات، على سبيل المثال، قد يكون الحافز أو الدافع هو مجموعة من الناس، مثل مندوبي مبيعات السيارات. ويتبع تحديد الدوافع والحوافز المستهدفة قائمة بأزواج الصفات. يمثل كل زوج طرفي نقيض من سلسلة متصلة، محددة بالصفات (على سبيل المثال، أمين أو صادق وغير أمين). كما هو موضح في المثال أدناه، هناك عدة نقاط (سطور) بين الصفات التي تشكل خيارات الاستجابة:

مندوب مبيعات السيارات:								
أمين (صادق)	-	-	-	-	-	-	-	غير أمين
هادئ (لطيف)	-	-	-	-	-	-	-	مزعج

في الأساس، تمثل الخطوط الفردية (سبعة وتسعة أرقام مشتركة) نقاطاً على طول السلسلة المتصلة المحددة بواسطة الصفات. يضع المجيب علامة أو إشارة على أحد الخطوط للإشارة إلى النقطة الموجودة على طول السلسلة المتصلة التي يتسم بها تقييمه للحافز أو الدافع. على سبيل المثال، إذا كان شخص ما يعتبر مندوب مبيعات السيارات غير أمين للغاية، فقد يختار هو أو هي الخط الأقرب إلى تلك الصفة. يمكن التعبير عن وجهات النظر والآراء المتطرفة أو المعتدلة من خلال اختيار الخط الذي يجب وضع علامة عليه. بعد تقييم الدوافع “الحوافز” فيما يتعلق بزواج الصفة الأول، ينتقل الشخص إلى أزواج صفة إضافية مفصولة بخطوط.

إن الصفات التي يختارها الشخص قد تكون إما ثنائية القطب أو أحادية القطب، اعتماداً، كما هو الحال دائماً، على المنطق الذي تستند إليه الأسئلة البحثية التي يهدف المقياس إلى معالجتها. كل الصفات ثنائية القطب تعبر عن وجود سمات متناقضة، مثل الودية friendly والعدائية hostile. وتشير أزواج الصفات أحادية القطب إلى وجود سمة واحدة وعدم وجودها (غيابها)، مثل الودية friendly وغير الودية.

كما هو الحال مع مقياس “ليكرت”، يمكن أن يكون شكل وصيغة الاستجابة التفاضلية الدلالية متوافقاً إلى حد كبير مع النماذج النظرية التي تم عرضها في الفصول السابقة من هذا الكتاب. يمكن كتابة مجموعات من البنود للاستفادة من نفس المتغير الأساسي. على سبيل المثال، يمكن إضافة البنود التي تستخدم نقاط النهاية التالية، جدير بالثقة / غير جدير بالثقة، وعادل / وغير عادل، وصادق / غير صادق كنقاط نهاية إلى العبارة الأولى في المثال السابق بحيث تشكل مقياس “الشرف أو الأمانة”. يمكن تصور مثل هذا المقياس على أنه مجموعة من البنود التي تشترك في متغير كامن مشترك هو (الشرف أو الأمانة) وتتوافق مع الافتراضات التي تمت مناقشتها في الفصل الثاني. وبناءً على ذلك، يمكن إضافة وتحليل العشرات من البنود كل على حدة «لمقياس الأمانة» على النحو المبين في قسم لاحق فيما يتعلق بتقييم البنود.

#### التناظر البصري:

هناك تنسيق آخر لشكل وصيغة البند يشبه إلى حد ما القياس التفاضلي الدلالي وهو مقياس التناظر البصري. يقدم شكل وصيغة الاستجابة هذا للمستجيب خطأ متصلاً بين زوج من التوصيفات الذي يمثل نهايات على طرفي نقيض (متعكسة) لسلسلة متصلة. ويُطلب من الشخص الذي يكمل البند وضع علامة عند نقطة على الخط تمثل رأيه أو خبرته أو معتقده أو أي شيء يتم قياسه. إن مقياس التناظر البصري، كما يوحي مصطلح التناظرية في الاسم، هو مقياس مستمر أو متصل. يقوم فيه الباحث بتحديد دقة التمايز في تعيين الدرجات إلى النقاط على المقياس. وقد تم في وقت سابق مناقشة بعض مزايا وعيوب شكل أو صيغة الاستجابة المستمرة. وهناك مسألة أو مشكلة إضافية لم تتم إثارته في ذلك الوقت تتعلق باختلافات المحتملة في تفسير المساحة المادية من حيث صلتها بالقيم على المتصل. قد لا تعني العلامة الموضوعية في نقطة معينة على طول الخط نفس الشيء بالنسبة لأشخاص مختلفين، حتى عندما يتم تمييز نقاط النهاية للخط بشكل مماثل لجميع المجيبين. فكر في مقياس تناظر بصري للألم مثل التالي:

لا ألم على الإطلاق ..... أسوأ ألم واجهته أو جربته.

هل تشير الاستجابة في منتصف المقياس إلى الألم الذي يصيب نصف الوقت، أو الألم المستمر بنصف الشدة المحتملة، أو شيء آخر بالكامل؟ جزء من المشكلة مع قياس الألم هو أنه

يمكن تقييمه على أبعاد متعددة، بما في ذلك التكرار، والشدة، والمدة. كما أن ذكريات أسوأ ألم قد تعرض له شخص ما من المرجح أن تكون مشوهة. والواقع أن المقارنات بين الأفراد تزداد تعقيداً بفعل حقيقة مفادها أن الأشخاص المختلفين ربما قد تعرضوا لمستويات مختلفة من «أسوأ ألم». وبطبيعة الحال، لا شك أن بعض هذه المشاكل تكمن في الظاهرة المستخدمة في هذا المثال - وهي الألم (انظر «كليف» 2000، Keefe)، للحصول على مناقشة ممتازة لقياس الألم) - وليس مع المقياس في حد ذاته. ومع ذلك، يمكن أن توجد مشكلة تعيين للقيم بطريقة غير اعتيادية (عشوائية) على طول مقياس تناظري بصري لظواهر أخرى أيضاً.

ومن المميزات الرئيسية لمقاييس التناظر البصري أنها قد تكون حساسة للغاية (ماير، 1978، Mayer)، وهذا من شأنه أن يجعلها مفيدة بشكل خاص في قياس الظواهر قبل وبعد وقوع بعض الأحداث المتداخلة، مثل التدخل أو المعالجة التجريبية، التي تمارس تأثيراً ضعيفاً نسبياً. على سبيل المثال، قد لا يؤدي التوبيخ الخفيف أو المتوسط في سياق المعالجة التجريبية إلى إحداث تحول في مقياس لاحترام الذات ذي تدريج خماسي النقاط. ولكن ربما قد يحدث تحول دقيق ولكنه منهجي أو منتظم يؤدي إلى انخفاض القيم على مقياس تناظري بصري بين الناس في حالة «التوبيخ» لهذه التجربة الافتراضية. وقد تكون الحساسية أكثر فائدة عند فحص ودراسة التغيرات على مر الزمن داخل نفس الفرد وليس بين الأفراد (ماير، 1978، Mayer). وقد يكون الأمر كذلك لأنه في الحالة الأولى، لا يوجد خطأ إضافي بسبب الاختلافات الخارجية بين الأفراد.

وهناك ميزة أخرى محتملة لمقاييس التناظر البصري عند تكرارها بمرور الوقت، وهي أنه من الصعب أو المستحيل على الأشخاص ترميز «تشفير» إجاباتهم السابقة بدقة. بالاستمرار في المثال الوارد في الفقرة السابقة، ربما لا يجد شخص ما صعوبة كبيرة في تذكر أي من الخيارات الخمسة المرقمة بالنسبة لبند تقدير الذات الذي كان قد اختاره أو اختارته من قبل استجابة لشكل أو صيغة متعددة الاستجابات مثل مقياس «ليكرت». إلا أنه ما لم يتم اختيار إحدى نقاط النهاية لمقياس تناظر بصري، سيكون من الصعب أن نتذكر بدقة المكان الذي تم فيه وضع علامة على طول خط غير مميز. وقد يكون ذلك مفيداً إذا كان الباحث يشعر بالقلق من أن المجيبين قد ينحازون إلى الظهور بمظهر متسق مع مرور الوقت. ومن المفترض أن يختار الأشخاص الذين لديهم دوافع ليكونوا متسقين نفس الاستجابة بعد التعرض لتدخل تجريبي كما كان الحال قبل التعرض لمثل هذا التدخل. غير أن شكل أو صيغة التناظر البصري يستبعد هذا الاحتمال. إذا كانت استجابات ما بعد



المعالجة التجريبية قد خرجت بشكل ثابت (أي في نفس الاتجاه عادةً) عن الاستجابة السابقة قبل المعالجة بالنسبة للموضوعات التجريبية وبشكل عشوائي لعناصر الضبط والتحكم، فإن اختيار مقياس تناظر بصري قد يكون ساهم في الكشف عن ظاهرة دقيقة كانت أساليب أخرى تخفيها.

غالبًا ما تستخدم مقاييس التناظر البصري كمقاييس أحادية البند. وهذا له عيب كبير يتمثل في استبعاد أي تحديد للاتساق الداخلي. ومع المقياس أحادي البند، فإنه لا يمكن تحديد الثبات إلا من خلال استخدام طريقة إعادة الاختبار test-retest الموضحة في الفصل الثالث أو بالمقارنة مع مقاييس أخرى لنفس السمة التي لها خصائص سيكومترية راسخة. وتعاني الطريقة الأولى من مشكلات تقييمات إعادة الاختبار التي تمت مناقشتها سابقًا، ولا سيما استحالة التمييز بين عدم الاستقرار في عملية القياس وعدم استقرار الظاهرة التي يتم قياسها. أما الطريقة الأخيرة فهي في الواقع تقابل (تشبه) صدق المفهوم البنائي. ومع ذلك، نظرًا لأن الثبات شرط ضروري للصدق، يمكن للمرء أن يستنتج الثبات إذا كان هناك دليل على الصدق. ومع ذلك، قد تكون الإستراتيجية الأفضل هي تطوير عدة بنود تناظرية بصرية بحيث يمكن تحديد الاتساق الداخلي.

#### أشكال الاستجابة العددية والعمليات العصبية الأساسية:

تشير دراسة أجرتها «زورزي وبرفيتليس وأوميليتا» Zorzi, Priftis, and Umilitá (2002) نُشرت في مجلة الطبيعة Nature إلى أن بعض خيارات الاستجابة قد تتوافق مع كيفية معالجة الدماغ للمعلومات الرقمية أو العددية. وفقاً لما ذكره هؤلاء المؤلفون، فإن الأرقام المرتبة في تسلسل، كما هو الحال مع مقياس "ليكرت" النموذجي، تعبر عن الكمية ليس فقط في قيمها الرقمية أو العددية ولكن في مواقعها. ويقترح هؤلاء المؤلفون أن الخط البصري للأرقام ليس مجرد تمثيل مناسب بل إنه يتوافق مع العمليات العصبية الأساسية. ولاحظوا أن الأشخاص الذين يعانون من إصابات مختلفة في الدماغ والتي تؤثر على الإدراك المكاني في المجال البصري يقومون بارتكاب أخطاء منهجية في مسائل الرياضيات البسيطة التي يتم تقديمها بصرياً (أو بصورة مرئية). ويرتبط الشذوذ "الانحراف" المكاني ونوع الأخطاء ارتباطاً وثيقاً. فالأفراد الذين لم يتمكنوا من إدراك المجال البصري أو المرئي الأيسر، عندما طُلب منهم الإشارة إلى نقطة الوسط أو المنتصف بين قيمتين معروضتين في ترتيب أو مصفوفة خطية، كانوا يخطئون دوماً "إلى اليمين". على سبيل المثال، عندما سئل الأفراد عن النقطة المتوسطة بين النقطتين "3" و"9"، تحولت الأخطاء

إلى اليمين (أي إلى القيم الأعلى). وبعكس المقياس من القيم الأعلى إلى القيم الأدنى استمر في إحداث تحولات إلى اليمين (الآن، القيم الأدنى). وعندما تم تقديم نفس المهام في شكل غير مرئي (على سبيل المثال، من خلال طرح السؤال عن متوسط 3 و9)، لم يظهر النمط. والواقع، لم يظهر هؤلاء الأفراد أي عجز في أداء الحسابات عندما لم يتم عرضها بصرياً. لم يظهر الأشخاص الخاضعين للضبط دون وجود الشدوذ أو الانحراف البصري نمط التحول لدى الذين يعانون من إصابات أو آفات الدماغ. ويخلص المؤلفون إلى أن عملهم يشكل «دليلاً قوياً على أن خط الأرقام الذهنية هو أكثر من مجرد تعبير مجازي أو تشبيه وأن «التفكير في الأرقام من الناحية المكانية (كما ذكر علماء الرياضيات العظماء) قد يكون أكثر كفاءة لأنه يقوم على أساس التمثيل العصبي الفعلي للأرقام» (ص 138). وعلى الرغم من أن هذه الدراسة، في حد ذاتها، قد لا تبرر استنتاجات قاطعة وسريعة، إلا أنها توفر أدلة أولية جديرة بالاهتمام على أن تقييم سلسلة خطية من الأرقام قد تتوافق مع الآليات العصبية الأساسية المتضمنة في التقييم الكمي. وإذا كان هذا هو الحال بالفعل، فإن خيارات الاستجابة المعروضة كصف من الأرقام قد تكون لها ميزة خاصة.

#### الخيارات الثنائية:

يوفر شكل وصيغة استجابة شائع آخر للأشخاص فرصة الاختيار بين الخيارات الثنائية لكل بند. وكانت الأمثلة السابقة لمقياس «ثرستون» Thurstone ومقياس «جتمان» Guttman تستخدم الخيارات الثنائية («موافق» و«غير موافق»)، على الرغم من أن المقاييس التي تحتوي على البنود متساوية الوزن قد تشتمل أيضاً على خيارات استجابة ثنائية. على سبيل المثال، قد يُطلب من الأشخاص التأشير على جميع الصفات الموجودة في قائمة يعتقدون أنها تنطبق على أنفسهم، أو قد يُطلب منهم الإجابة بـ «نعم» أو «لا» على قائمة ردود الفعل العاطفية التي قد تكون لديهم في بعض المواقف المحددة. وفي كلتا الحالتين، يمكن جمع أو دمج الاستجابات التي تعكس بنود تشترك في متغير كامن مشترك (على سبيل المثال، الصفات مثل «حزين» و«غير سعيد» و«الأزرق» والتي تمثل الاكتئاب) في درجة واحدة لهذا المفهوم البنائي.

ويتمثل أحد أوجه القصور الرئيسية في الاستجابات الثنائية في أن كل بند لا يمكن أن يكون له سوى أقل قدر من التباين (القابلية للتغير). وبالمثل، فإن أي زوج من البنود يمكن أن يكون له مستوى واحد فقط من مستويين من التباين (التباين المشترك): الاتفاق أو عدم الاتفاق. تذكر من

الفصل الثالث أن تباين المقياس الذي يتكون من عدة بنود متساوية الوزن أو الترجيح يساوي تماماً مجموع كافة العناصر في مصفوفة التباين لكل بند على حدة. وفي حالة البنود ثنائية الاستجابة، فإن كل بند يساهم بقدر ضئيل في هذا المجموع بسبب القيود في التباينات والتغيرات (التباينات المشتركة) المحتملة. والنتيجة العملية لذلك هي وجود حاجة إلى المزيد من البنود للحصول على نفس درجة تباين المقياس إذا كانت البنود ثنائية الاستجابة. ومع ذلك، عادةً ما تكون الإجابة على البنود ثنائية الاستجابة سهلة للغاية. ولذلك، فإن العبء الملقى على عاتق الشخص يكون منخفضاً بالنسبة لأي بند من البنود. على سبيل المثال، يمكن لمعظم الناس أن يقرروا بسرعة ما إذا كانت بعض الصفات هي أوصاف مناسبة لأنفسهم. ونتيجة لذلك، غالباً ما يكون الأشخاص على استعداد لإكمال المزيد من البنود ثنائية الاستجابة أكثر من تلك البنود التي تستخدم شكل أو صيغة تتطلب التركيز على الفروق الدقيقة. وبالتالي، فإن الشكل ثنائي الاستجابة قد يسمح للباحث بتحقيق تباين كافٍ في درجات المقياس عن طريق تجميع المعلومات على المزيد من البنود.

#### الأنطر الزمنية للبناء:

ثمة مشكلة أخرى تتعلق بشكل وتنسيق البنود تتمثل في الإطار الزمني المحدد أو الضمني. يناقش كيلي ومكجراث (Kelly and McGrath (1988)، في مجلد آخر من هذه السلسلة، أهمية النظر والأخذ في الاعتبار السمات الزمنية للمقاييس المختلفة. ولن تشير بعض المقاييس إلى إطار زمني، مما يعني وجود منظور زمني عالمي. على سبيل المثال، كثيراً ما تتضمن مقاييس موضع التحكم والسيطرة، على بنود تنطوي على إيمان دائم بالعلاقات السببية. وتفترض بعض البنود مثل «إذا اتخذت الإجراءات الصحيحة، فبوسعي أن أبقى في صحة جيدة» (والستون وآخرون (Wallston et al., 1978) أن هذا الاعتقاد مستقر نسبياً. ويتوافق هذا مع التوصيف النظري لموضع التحكم والسيطرة باعتباره متوسطاً عاماً وليس محدداً لضبط النتائج (على الرغم من حدوث تحول نحو مزيد من الخصوصية والتحديد في مقاييس لاحقة لموضع معتقدات التحكم - مثل ديفاليس وآخرون، 1985 (DeVellis et al., 1985)). وهناك مقاييس أخرى لتقييم الظواهر المؤقتة (العابرة) نسبياً. فالإكتئاب، على سبيل المثال، يمكن أن يختلف بمرور الوقت، وقد أقرت المقاييس التي تقيسه بهذه النقطة (ماير Mayer, 1978). على سبيل المثال، يستخدم مقياس الإكتئاب الخاص بمركز الدراسات الوبائية (رادلوف Radloff, 1977) الذي يستخدم على نطاق واسع تنسيقاً يطلب من

المستجيبين الإشارة إلى عدد المرات التي عانوا فيها حالات مزاجية مختلفة خلال الأسبوع الماضي. وبعض المقاييس، مثل مقاييس القلق (1970 Lushene, & Gorsuch, Spielberger, e.g.) تم تطويرها بأشكال مختلفة وتهدف إلى تقييم الحالات العابرة (المؤقتة) نسبياً أو السمات الدائمة نسبياً (Zuckerman, 1983).

وينبغي على الباحث أن يختار إطاراً زمنياً لمقياس نشط (إيجابي) وليس بشكل سلبي. وتُعد النظرية بمثابة دليل مهم لهذه العملية. هل تشكل الظاهرة موضع الاهتمام جانباً أساسياً ودائماً لشخصيات الأفراد، أم أنها من المرجح أن تعتمد على الظروف المتغيرة؟ هل المقصود من المقياس هو الكشف عن الاختلافات الدقيقة التي تحدث خلال فترة زمنية قصيرة (على سبيل المثال، زيادة التأثير السلبي بعد مشاهدة فيلم حزين) أو التغيرات التي قد تتطور على مدى العمر (على سبيل المثال، السياسات المحافظة المتصاعدة مع زيادة السن)؟

وفي الختام، ينبغي أن تعكس أشكال وتنسيقات البنود، بما في ذلك خيارات الاستجابة والتعليمات، طبيعة المتغير الكامن محل الاهتمام والاستخدامات المستهدفة للمقياس.

#### الخطوة الرابعة: مراجعة تجمع البنود الأولية من قبل الخبراء:

حتى الآن، ناقشنا الحاجة إلى تحديد وتوضيح الظاهرة محل الاهتمام تحديداً واضحاً، وتوليد تكوين مجموعة من البنود المناسبة، واختيار شكل الاستجابة أو الردود لتلك البنود. الخطوة التالية في هذه العملية هي قيام مجموعة من الأشخاص الذين لديهم دراية ومعرفة في مجال المحتوى بمراجعة تجمع البنود. وتحقق هذه المراجعة أغراضاً متعددة تتعلق بتعظيم صلاحية المحتوى للمقياس (انظر الفصل الرابع).

أولاً، قد تؤدي مراجعة الخبراء لتجمع البنود الخاص بك إلى تأكيد أو إبطال التعريف الخاص بالظاهرة. ويمكنك أن تطلب من فريق الخبراء (على سبيل المثال، الزملاء الذين عملوا بشكل مكثف ومستفيض مع المفهوم البنائي محل البحث أو الظواهر ذات الصلة) تقييم مدى ارتباط وأهمية كل بند من البنود بما تنوي قياسه. ويكون هذا مفيداً بشكل خاص إذا كنت تقوم بتطوير مقياس يتكون من مقاييس منفصلة لقياس مفاهيم بنائية متعددة. إذا كانت لديك عناية كبيرة عند وضعك وتطويرك للبنود، فإن الخبراء لن يجدوا صعوبة كبيرة في تحديد البنود التي تتوافق مع

المفاهيم البنائية. في الأساس، فإن أفكارك حول ما يقيسه كل بند تمثل الفرضية، وردود وإجابات الخبراء هي البيانات التي تؤكد أو تنفي الفرضية (بمعنى إنها تمثل اختبار هذه الفرضية). حتى إذا كان الغرض من كافة البنود مخصصاً لقياس سمة واحدة أو مفهوم بنائي واحد، فإن مراجعة الخبراء تكون مفيدة. إذا قرأ الخبراء شيئاً ما في بند لم تكن تنوي لإدراجه وتضمينه، فقد يفعل الأشخاص الذين يكملون مقياساً نهائياً نفس الشيء.

عادةً ما تتضمن آليات الحصول على تقييمات تتعلق بصلة البند على تزويد فريق الخبراء بتعريفك العملي للمفهوم البنائي، ثم يُطلب منهم تقييم كل بند من حيث أهميته وصلته بالنسبة للمفهوم البنائي كما قمت بتحديدته وتعريفه. وقد يستلزم هذا الأمر مجرد تصنيف أهمية أو صلة كل بند على أنها عالية أو متوسطة أو منخفضة. بالإضافة إلى ذلك، قد تدعو الخبراء إلى التعليق على البنود كل على حدة على النحو الذي يروونه مناسباً، وهذا يجعل عملهم أكثر صعوبة إلى حد ما، ولكنه يمكن أن يوفر معلومات ممتازة. وقد تمنحك بعض التعليقات الثاقبة أو المتبصرة حول سبب غموض بعض البنود، على سبيل المثال، منظوراً جديداً عن كيفية محاولة قياس مفهومك البنائي.

كما يمكن للمراجعين أيضاً تقييم وضوح البنود ومدى إيجازها ودقتها. قد يكون محتوى أحد البنود ذا صلة بالمفهوم البنائي، ولكن صياغته قد تكون فيها إشكالية. وهذا يؤثر على ثبات البند، لأن وجود بند غامض أو غير واضح، بدرجة أكبر من بند واضح، يمكن أن يعكس عوامل خارجة عن المتغير الكامن. في التعليمات الخاصة التي تقدمها إلى المراجعين، اطلب منهم الإشارة إلى البنود غير الملائمة أو المربكة واقتراح صيغ بديلة لها، إذا كانوا يميلون إلى ذلك.

وتتمثل الخدمة الثالثة التي يمكن أن يقدمها خبراء المراجعة في الإشارة إلى طرق الاستفادة من الظاهرة التي فشلت في إدراجها. قد تكون هناك مقاربة كاملة قد أغفلتها وتجاهلتها. على سبيل المثال، ربما تكون قد قمت بإدراج العديد من البنود التي تشير إلى المرض في مجموعة من البنود المعنية بالمعتقدات الصحية ولكنها لم تعتبر الجروح أو الإصابات بمثابة انحرافٍ آخر ذي صلة عن الصحة. من خلال مراجعة الطرق المتنوعة التي قمت بها لتجسيد والتعبير بها عن الظاهرة موضع الاهتمام، يمكن أن يساعدك المراجعون في زيادة صلاحية المحتوى للمقياس الخاص بك إلى أقصى حد.

كلمة تحذير أخيرة بشأن رأي الخبراء: إن القرار النهائي بقبول نصيحة الخبراء أو رفضها هو مسؤوليتك كمطور للمقياس. وفي بعض الأحيان، قد لا يفهم خبراء المحتوى مبادئ بناء المقياس، وقد يؤدي هذا إلى نصيحة سيئة. ومن التوصيات التي واجهها كثير من الزملاء الذين لا يتمتعون بخبرة كبيرة في مجال تطوير المقياس هي التخلص من البنود التي تتعلق بنفس الشيء. وكما سبقتنا مناقشته، فإن حذف جميع حالات التكرار من تجمع البنود أو المقياس النهائي سيكون خطأ فادحاً لأن التكرار هو جزء لا يتجزأ من الاتساق الداخلي. ومع ذلك، قد يشير هذا التعليق إلى أن صياغة البنود، ومفرداتها وتركيبها اللغوي متشابهة للغاية ويمكن تحسينها. انتبه جيداً لكل الاقتراحات التي تتلقاها من خبراء المحتوى، ثم اتخذ قرارك عن علم وبصيرة حول كيفية استخدام نصائحهم.

وفي هذه المرحلة من العملية، يكون لدى مطور المقياس مجموعة من البنود التي تمت مراجعتها من قبل الخبراء وتعديلها وفقاً لذلك. وقد حان الوقت الآن للتقدم إلى الخطوة التالية.

#### الخطوة الخامسة: النظر في إدراج وتضمين البنود التي تحقق الصدق:

من الواضح أن جوهر استبيان تطوير المقياس هو مجموعة البنود التي سينبثق منها المقياس قيد التطوير. وعلى أية حال، فإن التبصر وبعض الحكمة يمكن أن تؤدي ثمارها بشكل رائع. وقد يكون من الممكن والملائم نسبياً إدراج وتضمين بعض البنود الإضافية في الاستبيان نفسه التي ستساعد في تحديد صدق المقياس النهائي. هناك نوعان على الأقل من البنود التي ينبغي مراعاتها وأخذها في الاعتبار.

النوع الأول من البنود الذي قد يختار مطور المقياس إدراجه وتضمينه في استبيان يعمل أو يساعد على اكتشاف العيوب أو المشكلات. قد لا يجيب المستجيبون على البنود ذات الأهمية الأساسية للأسباب التي تفترضها، وقد تكون هناك دوافع أخرى تؤثر على استجاباتهم وردودهم. إن العلم بذلك ومعرفة في وقت مبكر يُعد أمراً مفيداً. ومن بين أنواع الدوافع التي يمكن تقييمها بسهولة إلى حد ما هو المرغوبة الاجتماعية desirability social. وإذا كان الفرد متحمساً بقوة لتقديم نفسه أو نفسها بطريقة يعتبرها المجتمع إيجابية، فقد تكون الاستجابات على البنود مشوهة أو محرفة. ويسمح تضمين مقياس المرغوبة الاجتماعية للباحث بتقييم مدى تأثر كل بند من البنود على حدة بشدة بالمرغوبة الاجتماعية. وينبغي اعتبار البنود التي ترتبط ارتباطاً كبيراً بدرجة

المرغوبة الاجتماعية التي يتم الحصول عليها مرشحة للاستبعاد ما لم يكن هناك سبب نظري سليم يشير إلى خلاف ذلك. وقد تم وضع وتطوير مقياس موجز ومفيد للمرغوبة الاجتماعية من قبل "ستراهان وجيرباسي" (1972) Strahan and Gerbasi. ويمكن إدراج هذا المقياس المكون من عشرة بنود بشكل ملائم في استبيان.

وهناك مصادر أخرى لبنود الكشف عن اتجاهات الاستجابة غير المرغوب فيها "أناستاسي" (Anastasi, 1968). يتضمن اختبار مينيسوتا المتعدد الأوجه للشخصية The Minnesota Multiphasic Personality Inventory (هاثاواي وماكنلي، Hathaway & McKinley, 1967؛ هاثاواي وميهل، 1951) (Hathaway & Meehl, 1951) العديد من المقاييس التي تهدف إلى الكشف عن أوجه التحيز المختلفة في الاستجابة. وفي بعض الحالات، قد يكون من المناسب إدراج هذه الأنواع من المقاييس.

أما الفئة الأخرى من البنود التي ينبغي أخذها في الاعتبار لإدراجها في هذه المرحلة فهي تتعلق بصدق المفهوم البنائي للمقياس. ووفقاً لما تمت مناقشته في الفصل الرابع، إذا كانت النظرية تؤكد أن الظاهرة التي تقوم بقياسها ترتبط بمفاهيم بنائية أخرى، فإن أداء المقياس في مقابل مقاييس تلك المفاهيم البنائية الأخرى يمكن أن يكون دليلاً على صدقه. وبدلاً من بذل جهد منفصل للتحقق من الصدق بعد تكوين المقياس النهائي، قد يكون من الممكن إدراج مقاييس المفاهيم البنائية ذات الصلة في هذه المرحلة. ويمكن أن يوفر نمط العلاقات الناتجة دعماً للدعوات المتعلقة بالصدق أو بدلاً من ذلك، يقدم أدلة إذا لم يكن أداء مجموعة البنود كما كان متوقعاً.

#### الخطوة السادسة: تطبيق البنود على عينة تطوير المقياس:

بعد تحديد البنود المرتبطة بالمفهوم البنائي والصدق التي يجب تضمينها وإدراجها في الاستبانة الخاصة بك، يجب عليك تطبيقها، جنباً إلى جنب مع مجموعة البنود الجديدة، على بعض الأشخاص. يجب أن تكون عينة الأشخاص كبيرة. كم يبلغ حجمها الكبير؟ من الصعب التوصل إلى توافق في الآراء بشأن هذه المسألة. دعونا ندرس الأساس المنطقي لعينة كبيرة. ويشير «نونالي» (Nunnally, 1978) إلى أن مسألة أخذ العينات الأولية في تطوير المقياس تنطوي على أخذ عينات من البنود من عالم افتراضي (راجع غيسيلي، كامبل، وزيدك، Ghiselli, Campbell, & Zedeck, 1981). ومن أجل التركيز على مدى كفاية البنود، ينبغي أن تكون العينة كبيرة بما فيه

الكفاية للقضاء على تباين الأشخاص باعتباره مصدر قلق كبير. ويشير «نونالي Nunnally» إلى أن 300 شخص هم عدد كافٍ. ومع ذلك، فإن التجربة العملية تشير إلى أن المقاييس قد تم وضعها وتطويرها بنجاح باستخدام عينات أصغر حجماً. كما أن عدد البنود وعدد المقاييس التي سيتم استخراجها واستخلاصها له تأثير أيضاً على مسألة حجم العينة. وإذا كان الهدف استخراج مقياس واحد فقط من مجموعة تضم حوالي 20 بنداً، فإن عينة حجمها أقل من 300 شخص قد تكون كافية.

هناك العديد من المخاطر في استخدام عدد قليل جداً من الأشخاص. أولاً، قد لا تكون أنماط التباين بين البنود مستقرًا. قد يتحول بند ما يبدو أنه يزيد من الاتساق الداخلي إلى أن يكون عديم القيمة عند استخدامه في عينة منفصلة. إذا تم اختيار البنود لإدراجها (كما هي بشكل جيد للغاية) على أساس مساهمتها في معامل “ألفا”، فإن وجود عينة تطويرية صغيرة يمكن أن ترسم صورة وردية غير دقيقة للاتساق الداخلي. وعندما تكون نسبة الأشخاص إلى البنود منخفضة نسبياً وحجم العينة غير كبير، يمكن أن تتأثر الارتباطات المتبادلة بين البنود بالصدفة chance بدرجة كبيرة إلى حد ما. وعندما تتم إعادة إدارة (تطبيق) المقياس الذي تم اختيار بنوده في ظل هذه الظروف، فإن العوامل المحتملة التي جعلت بعض البنود تبدو جيدة في البداية لم تعد صالحة. وبالتالي، فإن معامل “ألفا” الذي يتم الحصول عليه في مناسبات «دراسات» أخرى غير دراسة التطوير الأولية للمقياس قد تكون أقل من المتوقع. وعلى نحو مماثل، قد يتم استبعاد أي بند من المحتمل أن يكون جيداً لأن ارتباطه بينود أخرى كانت منخفضة بمحض الصدفة البحتة.

وثمة مشكلة ثانية محتملة لحجم العينة الصغير هي أن عينة التطوير قد لا تمثل المجتمع الإحصائي الذي يستهدفه المقياس. وبطبيعة الحال، قد يكون هذا هو الحال أيضاً إذا كانت عينة التطوير كبيرة، ولكن من المرجح أن تستبعد عينة صغيرة أنواعاً معينة من الأفراد. وبالتالي، يجب على مطور المقياس أن يأخذ في الاعتبار كلاً من حجم وتكوين عينة التطوير. قد يختار الباحث الدقيق معالجة إمكانية أو قابلية تعميم المقياس عبر المجتمعات الإحصائية (أو بعض الجوانب facet الأخرى) باستخدام دراسة إمكانية التعميم G-study، على النحو الذي تم توضيحه في الفصل الثالث.

ليست كل أنواع عدم تمثيل المجتمع متطابقة. هناك على الأقل طريقتان مختلفتان قد لا تمثل العينة من خلالهما المجتمع الإحصائي الأكبر. الأولى تتضمن مستوى السمة الموجودة في العينة



مقابل مستواها في المجتمع الإحصائي المستهدف. فعلى سبيل المثال، قد تمثل العينة نطاقاً أضيق من السمة عما كان متوقعاً في المجتمع الإحصائي. قد يكون هذا التضيق في النطاق غير متماثل أيضاً بحيث يكون متوسط الدرجات التي يتم الحصول عليها على مقياس العينة أعلى أو أقل بكثير مما يتوقعه الفرد بالنسبة للمجتمع الإحصائي. قد تختلف الآراء المتعلقة بالعمر القانوني الملائم لشرب الخمر في حرم الجامعة، على سبيل المثال، عن الآراء المتعلقة بالموضوع نفسه في المجتمع المحلي ككل. ولا تعني القيمة المتوسطة للسمة التي لا تكون ممثلة للمجتمع بالضرورة عدم أهلية العينة لأغراض تطوير المقياس. وقد يؤدي ذلك إلى توقعات غير دقيقة بشأن متوسطات المقياس في حين أنها لا تزال تقدم صورة دقيقة عن ثبات الاتساق الداخلي الذي يتمتع به المقياس. فعلى سبيل المثال، قد تؤدي عينة من هذا النوع إلى استنتاجات صحيحة حول البنود الأكثر ارتباطاً مع بعضها البعض.

وهناك نوع أكثر إزعاجاً من عدم تمثيل العينة للمجتمع ينطوي على عينة تختلف نوعياً وليس كمياً عن المجتمع الإحصائي المستهدف. على وجه التحديد، فإن العينة التي قد تختلف فيها العلاقات بين البنود أو المفاهيم البنائية عن المجتمع الإحصائي هي سبب يدعو للقلق. إذا كانت العينة غير معتادة أو استثنائية تماماً، فقد يكون للبنود معنى مختلف بالنسبة لهم عن معناها بالنسبة للأشخاص بشكل عام. قد تعكس أنماط الارتباط بين البنود سمات غير عادية مشتركة بين أفراد العينة ولكنها نادرة الوجود في المجتمع الأوسع. وبعبارة أخرى، قد تكون مجموعات البنود المترابطة التي تنشأ (من التحليل العاملي، مثلاً) غير نمطية أو غير مألوفة. إذا تم ذكر ذلك بشكل رسمي بعض الشيء، فقد يكون الهيكل أو البناء السببي الأساسي الذي يربط المتغيرات بالدرجات الحقيقية مختلفاً إذا كانت العينة تختلف عن المجتمع الإحصائي من نواح هامة. ولنتأمل هنا بعض الأمثلة الواضحة إلى حد ما: إذا لم يفهم أعضاء العينة المختارة كلمة رئيسية تتكرر بين البنود وذات صلة بالمفهوم البنائي، فإن استجاباتهم عليها ربما تخبرنا بالقليل أو لا تخبرنا بأي شيء حول كيفية أداء المقياس في ظل ظروف مختلفة. فكلمة مريض "sick" تعني "مريض ill" في الولايات المتحدة، ولكنها تعني "الشعور بالغثيان nauseated" (أي الشعور بمرض في المعدة) في إنجلترا. قد يكون لمجموعة من الأسئلة حول المرض illness التي تم تطويرها لمجموعة ما معنى مختلف بشكل ملحوظ بالنسبة للمجموعة الأخرى. إذا كان المقياس يتعلق بمشكلة صحية معينة لا ترتبط عادةً بالغثيان (مثل التهاب المفاصل)، فقد تتجمع البنود التي تستخدم كلمة "مريض ill" معاً بسبب

معناها المميز إذا كانت العينة بريطانية. ومن ناحية أخرى، فإن العينة من الولايات المتحدة الأمريكية من غير المحتمل أن تميز أو تفرق بين العبارات المتعلقة بالمرض وغيره من البنود ذات الصلة بالصحة. حتى داخل الولايات المتحدة الأمريكية، يمكن أن تكون لنفس الكلمة معانٍ مختلفة. فبين سكان الجنوب في المناطق الريفية، على سبيل المثال، فإن تعبير “الدم الفاسد” يستخدم أحياناً كناية عن الأمراض التناسلية “مرض الزهري”، بينما يعني العداء في أجزاء أخرى من البلاد. وإذا كان هناك بند يناقش “الدم الفاسد بين الأقارب” وكان أدائه مختلفاً بين عينة من أهل الجنوب في المناطق الريفية مقابل عينات أخرى، فإن هذا لن يكون بالأمر المفاجئ أو المستغرب.

يمكن أن تؤدي عواقب هذا النوع الثاني من عدم تمثيل العينة للمجتمع إلى إلحاق ضرر بالغ بجهود تطوير المقياس. قد تكون البنية الأساسية التي تنشأ - وهي أنماط التغير بين البنود المهمة للغاية لقضايا ثبات المقياس - بمثابة اختلاف من العينة المستخدمة في التطوير. إذا كان لدى الباحث سبب للاعتقاد بأن المعنى المنسوب للبنود قد يكون غير شائع أو غير نمطي بين عينة التطوير، فينبغي توخي الحذر الشديد في تفسير النتائج التي يتم الحصول عليها من تلك العينة.

#### الخطوة السابعة: تقييم البنود:

وبعد تطوير المجموعة الأولية من البنود وتدقيق هذه البنود (تطبيقها) على عينة كبيرة وممثلة بشكل مناسب، فقد حان الوقت لتقييم أداء كل بند على حدة بحيث يمكن تحديد البنود المناسبة التي تشكل أو تكون المقياس. وهذا، من جوانب عديدة، هو جوهر أو أساس عملية تطوير المقياس. وربما يأتي تقييم البند في المرتبة الثانية فيما يتعلق بتطوير البنود من حيث أهميته فقط.

#### الفحص المبدئي لأداء البنود:

عند مناقشة تطوير البند، أشرنا إلى بعض الصفات المرغوب فيها التي ينبغي أن تكون في بنود المقياس. دعونا نعيد النظر في هذه القضية أو المسألة. إن الجودة النهائية التي نسعى إلى وجودها في أي بند هي أن يكون ارتباطه عالٍ مع الدرجة الحقيقية للمتغير الكامن. ويأتي هذا مباشرةً من مناقشة الثبات في الفصل الثالث. ولا يمكننا أن نقيم الدرجة الحقيقية بشكل مباشر (إذا استطعنا القيام بذلك، فربما لن نحتاج إلى مقياس)، وبالتالي، لا يمكننا أن نحسب ارتباطها المتبادل مع البنود بشكل مباشر. ومع ذلك، يمكننا إجراء استدلالات بناءً على نماذج القياس الرسمية التي

تمت مناقشتها حتى الآن. عند مناقشة الاختبارات المتكافئة في الفصل الثاني، أشرت إلى أن الارتباط بين أي بندين يساوي مربع الارتباط بين أي بند من البندين والدرجة الحقيقية. هذه القيمة التربيعية هي ثبات كل بند من البنود. لذلك؛ يمكننا التعرف على العلاقات مع الدرجات الحقيقية من خلال الارتباطات المتبادلة بين البنود. وكلما كانت الارتباطات بين البنود مرتفعة، كان ثبات كل بند من البنود على حدة مرتفعاً (أي، كلما كانت البنود مرتبطة ارتباطاً وثيقاً بالدرجة الحقيقية). وكلما كان كل بند من البنود على حدة أكثر ثباتاً، كان المقياس الذي يؤلفونه أكثر ثباتاً (على افتراض أنها تشترك في متغير كامن مشترك). لذلك؛ فإن أول صفة للجودة نسعى إليها في مجموعة من بنود المقياس هي أنها تكون مترابطة ترابطاً قوياً. وإحدى الطرق لتحديد كيفية أو مدى الارتباط المتبادل بين البنود تتمثل في فحص مصفوفة الارتباط.

### عكس تسجيل الدرجات:

إذا كانت هناك بنود ارتباطها ارتباط سلبي مع بنود أخرى، فينبغي عندئذ النظر في مدى مناسبة إجراء تسجيل درجات تلك البنود في صورة عكسية (أي التصحيح العكسي لتلك البنود). وفي وقت سابق، اقترحت أن البنود التي صيغت في اتجاهات متعاكسة يمكن أن تثير مشاكل. ولكن في بعض الأحيان، قد ينتهي بنا الأمر دون قصد إلى بنود مرتبطة مع بعضها البعض ارتباطاً سلبياً. وقد يحدث هذا، على سبيل المثال، إذا توقعنا في البداية مجموعتين منفصلتين من البنود (تتعلق بالسعادة والحزن، على سبيل المثال)، ولكننا قررنا لسبب ما أن يتم دمجها في مجموعة واحدة. ويمكننا بعد ذلك أن ينتهي بنا الأمر إلى عبارات ترتبط بشكل متساوٍ بالمفهوم البنائي المركب الجديد (على سبيل المثال، التأثير)، ولكن بعضها قد يكون إيجابياً وبعضها الآخر سلبياً. «أنا سعيد» و«أنا حزين» كلاهما يتعلق على حد سواء بالتأثير؛ ومع ذلك، فهما متضادان. إذا أردنا الحصول على درجات عالية على مقياسنا لقياس السعادة، فسيتمتعين علينا أن نعطي قيمة عالية لتأييد أو الموافقة على البند «سعيد» وإعطاء قيمة منخفضة لتأييد البند «حزين». وهذا يعني، أننا سنعكس تسجيل درجة بند الحزن. وفي بعض الأحيان، تتم إدارة أو تطبيق البنود بطريقة تكون معكوسة بالفعل. فعلى سبيل المثال، قد يُطلب من الأشخاص وضع دائرة حول قيم رقمية أعلى لبيان الاتفاق مع بند «سعيد» وقيم أقل للإشارة إلى تأييد بند «حزين». ومن بين الطرق للقيام بذلك أن تكون التوصيفات اللفظية لخيارات الاستجابة (على سبيل المثال، «لا أوافق بشدة»، «لا أوافق إلى حد ما»، وما إلى

ذلك) دائماً بالترتيب نفسه لكل البنود ولكن مع ربط الأرقام بها إما تصاعدياً أو تنازلياً، اعتماداً على البند:

١- أشعر بالحزن في كثير من الأحيان.

٦	٥	٤	٣	٢	١
غير موافق بشدة	غير موافق إلى حدٍ ما	غير موافق بشكل معتدل	موافق بشكل معتدل	موافق إلى حدٍ ما	موافق بشدة

٢- في كثير من الأوقات، أكون سعيداً.

١	٢	٣	٤	٥	٦
غير موافق بشدة	غير موافق إلى حدٍ ما	غير موافق بشكل معتدل	موافق بشكل معتدل	موافق إلى حدٍ ما	موافق بشدة

وقد تتسبب هذه العملية في إرباك الشخص المجيب. فقد يتجاهل الناس الكلمات بعد أن يدركوا أنها هي واحدة لكل البنود. ومع ذلك، ربما يكون من الأفضل تغيير ترتيب التوصيفات (على سبيل المثال، من “غير موافق بشدة” إلى “موافق بشدة” من اليسار إلى اليمين بالنسبة لبعض البنود وعكس اتجاه البنود الأخرى). وثمة خيار آخر يتمثل في أن تكون كل من الأوصاف اللفظية والأرقام المناظرة هي نفسها بالنسبة لكل البنود، ولكن يتم إدخال قيم مختلفة لبنود معينة عند ترميز البيانات. إن تغيير الدرجات لبنود معينة عند الترميز يكون أمراً مملاً ومن المحتمل أن يكون عرضة للأخطاء على حد سواء. لكل شخص، يجب أن يُعطى كل بند يتم تسجيل درجته بطريقة عكسية، الاهتمام الخاص الذي ينطوي عليه تسجيل الدرجات العكسية، وهذا من شأنه أن يخلق العديد من الفرص للأخطاء.

وأسهل طريقة لتسجيل الدرجات العكسية هي القيام بذلك إلكترونياً بمجرد إدخال البيانات إلى جهاز الحاسب الآلي. تستطيع بعض بيانات الحاسب الآلي التعامل مع كافة الدرجات العكسية لبيانات كل الأشخاص. إذا كانت خيارات الاستجابة تحتوي على قيم رقمية وكان التحويل المطلوب

هو عكس ترتيب هذه القيم، فيمكن استخدام صيغة (معادلة) بسيطة. على سبيل المثال، افترض أنه تم تسجيل مجموعة من بنود الحالة المزاجية التي تمت صياغتها أو تنسيقها باستخدام مقياس «ليكرت» من القيمة (١) إلى القيمة (٧)، والأرقام الأعلى تشير إلى الموافقة. افترض كذلك أنه، لتسهيل الفهم، تم استخدام نفس شكل أو صيغة الاستجابة لكل من بنود الحالة المزاجية الإيجابية وبنود الحالة المزاجية السلبية. ومع ذلك، إذا تم تعيين درجات عالية في حالة الموافقة على، وتأيد بنود الحالة المزاجية الإيجابية، فإن المقياس يكون في الأساس مقياس الحالة المزاجية الإيجابية. إن التأيد أي الموافقة على بند الحالة المزاجية الإيجابية يجب أن يؤدي إلى قيمة عالية، وتأيد بند الحالة المزاجية السلبية يجب أن يسفر عن قيمة منخفضة، وهذا هو ما يمكن الحصول عليه إذا تم تغيير درجات الاستجابات بالنسبة لجميع البنود المزاجية السلبية من الدرجة (٧) إلى الدرجة (١)، وتغير الدرجة (٦) إلى الدرجة (٢)، وهكذا دواليك. ويمكن تحقيق هذا النوع من التحويل عن طريق إنشاء درجة جديدة من الدرجة القديمة باستخدام الصيغة أو المعادلة التالية:

$$NEW = (J + 1) - OLD$$

حيث تشير (NEW) و(OLD) إلى الدرجات المحولة والدرجات الأصلية على الترتيب، ويشير (J) إلى الرقم الأصلي لخيارات الاستجابة. في المثال المقدم آنفاً، فإن (J) تساوي الرقم (٧)، و(1 + J) سيكون مساوياً للقيمة (٨). وطرح الدرجة (٧) من الدرجة (٨) سيسفر عن الدرجة (١)، وطرح الدرجة (٦) سيسفر عن الدرجة (٢)، وهكذا.

قد لا يمكن تصحيح بعض الارتباطات السلبية بين البنود عن طريق عكس درجات البنود. على سبيل المثال، قد يؤدي عكس درجة بند معين إلى إزالة بعض الارتباطات السلبية ولكنها تؤدي إلى إنشاء ارتباطات أخرى، ويشير هذا عادةً إلى أن بعض البنود لا تنتمي ببساطة لأنها لا ترتبط بشكل متسق ببنود أخرى. وينبغي التخلص من وحذف أي بند يرتبط إيجابياً مع بعض البنود ويرتبط سلبياً مع البعض الآخر في مجموعة متجانسة إذا لم يؤدي أي نمط من بنود الدرجات العكسية إلى إزالة وإلغاء الارتباطات السالبة.

ارتباطات البند بالمقياس:

إذا أردنا الوصول إلى مجموعة من البنود المترابطة ببعضها البعض بدرجة كبيرة، فيجب أن يرتبط كل بند على حدة جوهرياً بمجموعة البنود المتبقية. ويمكننا فحص هذه الخاصية لكل بند عن طريق حساب ارتباط البند بالمقياس الخاص به  $correlation\ item-scale$ . وهناك نوعان من ارتباط المقياس بالبند، هما: الارتباط المصحح للبند بالمقياس  $item-corrected$   $correlation\ scale$  والذي يربط بين البند الذي يجري تقييمه بكل بنود المقياس، باستثناء البند نفسه، في حين أن الارتباط غير المصحح للبند بالمقياس  $item-uncorrected$   $correlation\ scale$  يربط البند المعني بمجموعة البنود المرشحة بأكملها، بما في ذلك البند نفسه. وإذا كانت هناك عشرة بنود يجري النظر فيها من أجل تكوين مقياس، فإن الارتباط المصحح للبند بالمقياس لأي بند من البنود العشرة سيتألف من ارتباط هذا البند بالبنود التسعة الأخرى. وسيتألف الارتباط غير المصحح للبند بالمقياس من ارتباط هذا البند بكل البنود العشرة. ومن الناحية النظرية، فإن القيمة غير المصححة تخبرنا عن مدى أو كيفية تمثيل البند في المقياس بأكمله. وهذا مشابه، على سبيل المثال، للربط بين مجموعة فرعية واحدة من اختبار الذكاء (IQ) وبين الاختبار بأكمله لتحديد ما إذا كان المقياس الفرعي وكيلاً  $proxy$  مناسباً أم لا. ومع ذلك، وعلى الرغم من أن الارتباط الكلي غير المصحح بين البنود  $uncorrected\ item-total\ correlation$  منطقي من الناحية المفاهيمية أو النظرية، فإن الواقع هو أن إدراج البند في المقياس من الممكن أن يؤدي إلى تضخم معامل الارتباط. وكلما قل عدد البنود في المجموعة، ازداد الفارق الذي سيحدثه إدراج أو استبعاد البند قيد الفحص والتدقيق. وبصفة عامة، ربما يكون من المناسب أو المفضل فحص الارتباط المصحح بين البند ومجموع البنود  $correlation\ item-total\ corrected$ . ويعتبر البند ذو القيمة العالية لهذا الارتباط المتبادل أكثر جاذبية (مرغوباً أكثر) من البند ذي القيمة المنخفضة.

#### تباينات البنود:

هناك خاصية أخرى ذات قيمة لبند المقياس وهي التباين العالي نسبياً. لنأخذ حالة متطرفة، إذا قام جميع الأفراد بالإجابة على بند معين بشكل متطابق أو مماثل، فلن يميز المقياس على الإطلاق بين الأفراد ذوي المستويات المختلفة من المفهوم البنائي الذي يتم قياسه وسيكون تباينه مساوياً للقيمة «صفر». وعلى النقيض من ذلك، إذا كانت عينة تطوير المقياس متنوعة فيما يتعلق بالسمة محل الاهتمام، فينبغي أن يكون مدى الدرجات التي يتم الحصول عليها لأحد البنود متنوعاً

أيضًا. وهذا يعني وجود تباين كبير إلى حد ما. وبطبيعة الحال، لا يكون من المرغوب فيه زيادة التباين بإضافة مكون الخطأ.

وقد تكون مقارنة تباينات البنود مفيدة أيضًا، خاصةً إذا كان الهدف هو تطوير أداة تلبية وتفي بافتراضات معادلة التكافؤ الأساسي لـ «تاو» essential tau equivalence. وعلى الرغم من أن معادلة التكافؤ الأساسي لـ «تاو» لا تفترض وجود تباينات متساوية بين البنود، إلا أنها تفترض وجود تغيرات متساوية بين البنود التي لها درجة حقيقية. وبالتالي، عندما يتم الوفاء بافتراضات هذا النموذج، فإن الاختلافات أو الفروق في تباينات البنود تكون نتيجة إلى عدم الاتساق في تباينات أخطاء البنود. وكاعتبار عملي، قد يكون الفرق في تباينات البنود بشكل ملحوظ غير مرغوب فيها، لأنها قد تشير إما إلى مصادر خطأ مختلفة اختلافًا جوهريًا من بند إلى آخر أو فشل في تلبية شرط وجود تباينات مشتركة «تغيرات» متساوية بين البند والدرجة الحقيقية والتي يتطلبها نموذج التكافؤ الأساسي لـ «تاو». وبطبيعة الحال، لن ينطبق هذا إذا كان النموذج التوجيهي للتطوير هو النموذج المتجانس congeneric model، والذي لا يتطلب أي شيء غير البنود التي تشترك في متغير كامن مشترك.

#### متوسطات البنود:

من المرغوب فيه أيضًا أن يكون المتوسط قريباً من مركز المدى للدرجات المحتملة. على سبيل المثال، إذا كانت خيارات الاستجابة لكل بند تتراوح بين الدرجة “١” (المناظرة لخيار الاستجابة “لا أوافق بشدة”) إلى الدرجة “٧” (المناظرة لخيار الاستجابة “أوافق بشدة”), فإن البند الذي يكون متوسطه قريباً من الدرجة “٤” سيكون بنداً مثاليًا. وإذا كان المتوسط قريباً من أحد طرفي المدى للمقياس، فقد يفشل البند في اكتشاف قيم معينة من المفهوم البنائي. قد يشير تجمع أو تراكم الدرجات عند القيمة “٧”, على سبيل المثال، إلى أن البند لم يكن مصاغاً بقوة كافية (أي إنه من النادر أن تجد أي شخص قد لا يوافق عليه).

بشكل عام، فإن البنود ذات المتوسطات القريبة جداً من الحد الأقصى لمدى الاستجابة سيكون لها تباينات منخفضة، والبنود التي تختلف على نطاق ضيق سوف ترتبط ارتباطاً ضعيفاً أو شيئاً بالبنود الأخرى. وكما ذكرنا سابقاً، لا يمكن أن يكون البند الذي لا يتغير vary أن يتغير covary. وبالتالي، سوف يميل المتوسط غير المتوازن أو التباين المنخفض لأي سبب من الأسباب

إلى الحد من وتقليل ارتباط أي بند بالبند الأخرى. ونتيجة لذلك، يمكنك عادةً التركيز بشكل أساسي على نمط الارتباطات بين البنود كمقياس لقيمتها المحتملة. ومع ذلك، يُعد فحص المتوسطات والتباينات بمثابة فحص مزدوج مفيد بمجرد إجراء اختيار مبدئي للبنود على أساس الارتباطات المتبادلة.

### البعدية (الأبعاد):

إن مجموعة من البنود ليست بالضرورة أن تشكل مقياساً. قد لا يكون للبنود متغير أساسي مشترك (كما هو الحال في المؤشر index أو المتغير الناشئ) أو قد يكون لها عدة متغيرات. يُعد تحديد طبيعة المتغيرات الكامنة التي تقوم عليها مجموعة بنود أمراً بالغ الأهمية. على سبيل المثال، الافتراض الذي يستند عليه معامل “ألفا” هو أن مجموعة البنود تكون أحادية البعد. وأفضل وسيلة لتحديد أي مجموعات البنود، إن وجدت، تشكل مجموعة أحادية البعد هو التحليل العاملي. وهذا الموضوع مهم بما فيه الكفاية بحيث يستحق فصلاً كاملاً (انظر الفصل السادس). وعلى الرغم من أن التحليل العاملي يتطلب أحجام عينات كبيرة، فإن تطوير المقياس بشكل عام يتطلب أيضاً عينات ذات أحجام كبيرة. وإذا كان هناك عدد قليل جداً من المجيبين على التحليل العاملي، فإن عملية تطوير المقياس برمتها قد تتعرض للخطر. وبالتالي، ينبغي أن يكون التحليل العاملي من أي نوع بشكل عام جزءاً من عملية وضع وتطوير المقياس في هذه المرحلة.

### الثبات (الموثوقية):

من بين المؤشرات الأكثر أهمية لجودة المقياس هو معامل الثبات، ألفا. في الواقع، إن كل مشاكل البنود التي تمت مناقشتها حتى الآن - تتمثل في المتوسط اللامركزي، والتباين الضعيف، والارتباطات السلبية بين البنود، والارتباطات المنخفضة بين المقياس والبند، والارتباطات الضعيفة بين البنود وبعضها البعض - ستميل إلى تقليل معامل “ألفا” وربما تبرر استخدام مؤشر بديل، مثل معامل “أوميغا”. لذلك؛ بعد أن نختار البنود الخاصة بنا - نقوم بالتخلص من البنود الفقيرة والاحتفاظ بالبنود الجيدة - ويُعد معامل “ألفا” إحدى طرق تقييم مدى نجاحنا، على افتراض أن البنود تلبى وتفي بالافتراضات الخاصة باستخدام معامل “ألفا”. إن معامل “ألفا” هو مؤشر على نسبة التباين في درجات المقياس التي تُعزى إلى الدرجة الحقيقية. وهناك العديد من الخيارات لحساب معامل “ألفا”، تختلف في درجة الأتمتة (أي استخدام آلات الحوسبة). تحتوي بعض حزم



الحاسب الآلي على برامج لتحليل البنود التي تقوم بحساب معامل "ألفا". في الحزمة الإحصائية للعلوم الاجتماعية والمعروف اختصاراً ببرنامج SPSS، يحسب الإجراء RELIABILITY معامل "ألفا" للمقياس بالكامل ولجميع النسخ (1 - k) (أي لكل نسخة ممكنة مع حذف بند واحد). ويوفر البرنامج أيضاً الارتباطات المصححة وغير المصححة بين المقياس والبند. ويتضمن برنامج "نظام التحليل الإحصائي statistical analysis system" والمعروف اختصاراً ببرنامج SAS حسابات معامل "ألفا" كخاصية لإجراء الارتباط، والمعروفة بـ PROC CORR. من خلال إدراج الخيار "ألفا" ALPHA في بيان PROC CORR، سيتم التعامل مع المتغيرات المدرجة في بيان VAR على أنها مقياس وسيتم حساب معامل "ألفا" للمجموعة الكاملة من البنود وكذلك جميع مجموعات البنود (1 - k) الممكنة. كما يوفر أيضاً الارتباطات بين المقياس والبند.

وهناك خيار آخر لحساب معامل "ألفا" هو القيام بذلك يدوياً. في حالة توفر تباينات لكل بند من البنود على حدة والمقياس ككل، فيمكن التعويض بهم في المعادلة أو الصيغة الأولى لحساب معامل "ألفا" التي تمت مناقشتها في الفصل الثالث. أو يمكن استخدام صيغة أو معادلة "سبيرمان - براون" Spearman-Brown، والتي تم تقديمها أيضاً في الفصل الثالث. وتستخدم هذه المعادلة المعلومات المتوفرة من مصفوفة ارتباط بدلاً من التباينات كأساس لحساب معامل "ألفا". ويتمثل أحد أوجه القصور في هذا النهج أو الطريقة في أن الارتباطات عبارة عن تباينات موحدة (أي تمت معاييرها)، وقد تؤثر معايير البنود كل على حدة على قيمة معامل "ألفا". وإذا تم الالتزام بدقة بنموذج الاختبارات المتكافئة، فهذا أمر غير منطقي أو غير مهم لأن الارتباطات يفترض أن تكون متساوية. ومع ذلك، فهي تقريباً لا تكون متساوية تماماً. لا يتطلب نموذج الصورة الأساسية لمعادلة "تاو" وجود ارتباطات متساوية بين البنود، بل يتطلب فقط وجود تباينات متساوية. وبالتالي، فإن نسبة تباين كل بند على حدة والتي تُعزى إلى الخطأ لا تختلف في إطار هذا النموذج. ومع ذلك، ونظراً لأن صيغة "سبيرمان - براون" تعمل بالفعل مع متوسط الارتباطات المتبادلة بين البنود، وأن من بين الآثار المترتبة على نموذج الصورة الأساسية لمعادلة "تاو" أن متوسط ارتباطات المقياس بالبند يكون متساوياً لكل بند، فإنه لا توجد هناك مشكلة. ومع ذلك، يمكن أن تكون هناك فروق صغيرة (ولكن كبيرة في بعض الأحيان) بين قيم معامل "ألفا" التي يتم الحصول عليها من الطرق الحسابية المستندة إلى التباين مقابل الطرق الحسابية القائمة على الارتباط. ونظراً لأن

مصفوفة التباير تستخدم البيانات بشكل أنقى أو خام (بدون معايرة أي توحيد القياس)، فهي مفضلة ويجب استخدامها بشكل عام.

ومن الناحية النظرية، يمكن أن يأخذ معامل «ألفا» القيم من «صفر» إلى «الواحد الصحيح»، على الرغم من أنه من غير المرجح أن يحقق معامل «ألفا» أيًا من هذه القيم المتطرفة. إذا كانت قيمة معامل «ألفا» سالبة، فهناك خطأ ما. وتتمثل المشكلة المحتملة في الارتباطات السالبة (أو التبايرات) بين البنود. إذا حدث ذلك، فحاول عكس الدرجات عند تسجيلها reverse scoring أو حذف البنود كما هو موضح سابقًا في هذا الفصل. ويقترح «نونالي» (1978) Nunnally القيمة (٠,٧٠) كحد أدنى مقبول لمعامل «ألفا». ليس من غير المألوف رؤية المقاييس المنشورة ذات قيم منخفضة لمعامل «ألفا». يختلف علماء المنهجية والباحثون المختلفون بدأوا مرتبكين بشأن مستويات مختلفة من معامل «ألفا»، وأفضل بصفة شخصية لمدى أو قيم معامل «ألفا» للمقاييس البحثية على النحو التالي: أقل من (٠,٦٠) غير مقبول، بين (٠,٦٠) و(٠,٦٥) غير مرغوب، بين (٠,٦٥) و(٠,٧٠) الحد الأدنى المقبول، وبين (٠,٧٠) و(٠,٨٠) جيد، وبين (٠,٨٠) و(٠,٩٠) جيد جداً، وأكبر بكثير من (٠,٩٠) ينبغي للمرء أن يفكر في أن يقلل من طول المقياس (انظر القسم التالي). وأود أن أؤكد أن هذه المجموعات هي شخصية وذاتية لقيم معامل «ألفا». ولا أستطيع الدفاع عنها على أسس عقلانية بحتة. ومع ذلك، فإنها تعكس تجربتي وخبرتي ويبدو أنها تتداخل بشكل كبير مع تقييمات الباحثين الآخرين. تنطبق القيم التي اقترحتها على معاملات «ألفا» المستقرة. وأثناء التطوير، يتم اختيار البنود، إما بشكل مباشر أو غير مباشر، على أساس مساهمتها في معامل «ألفا». وقد يُعزى بعض التباير الظاهري بين البنود إلى الصدفة. لذلك؛ من المستحسن خلال مرحلة التطوير أن نسعى إلى الحصول على معاملات «ألفا» أعلى قليلاً مما ترغب فيه. وإذا ما تدهورت قيم معامل «ألفا» بعض الشيء عند استخدامها في سياق بحثي جديد، فإنها ستظل مرتفعة بشكل مقبول إلى حد ما. وكما ذكر سابقاً، إذا كانت عينة التطوير صغيرة، فيجب أن يكون الباحث قلقاً بشكل خاص من أن التقديرات الأولية لمعامل «ألفا» التي يتم الحصول عليها أثناء تطوير المقياس قد لا تكون مستقرة. وكما سنرى، فإن هذا هو الحال أيضاً عندما يكون عدد البنود التي تشكل المقياس صغيراً.

ومن المواقف أو الحالات التي لا تنطبق عليها «نطاقات الراحة» comfort ranges المقترحة لمعامل «ألفا» عندما يتم تطوير مقياس يتطلب دقة بالغة الأهمية. وتعتبر الحالات

السريرية مثالاً على ذلك. وتكون الإرشادات المقترحة «المبادئ التوجيهية» مناسبة للأدوات البحثية التي ستستخدم مع بيانات المجموعات. على سبيل المثال، من المحتمل أن يكون المقياس الذي تبلغ قيمة معامل «ألفا» له (0.85) مناسباً تماماً للاستخدام في دراسة تقارن بين المجموعات فيما يتعلق بالمفهوم البنائي الذي يتم قياسه. يتطلب التقييم الفردي، ولا سيما عندما تعتمد القرارات المهمة على هذا التقييم، مستوى أعلى من ذلك بكثير. إن المقاييس المخصصة للتشخيص الفردي، أو التوظيف، أو التوظيف الأكاديمي، أو غير ذلك من الأغراض المهمة ينبغي أن تكون لها درجة أعلى من الثبات، في منتصف التسعينيات، على سبيل المثال.

في بعض الحالات، مثل الحالات التي يتكون فيها المقياس من بند واحد، سيكون من المستحيل استخدام معامل «ألفا» كمؤشر للثبات. وإذا كان ذلك ممكناً، فيجب إجراء تقييم معين للثبات. وقد يكون ارتباط إعادة الاختبار Test-retest correlation هو الخيار الوحيد في حالة المقياس المكون من بند واحد. وعلى الرغم من أن مؤشر الثبات هذا غير كامل، كما هو موضح في الفصل الثالث، فإنه من الواضح أنه أفضل من عدم تقييم الثبات على الإطلاق. والبديل المفضل، إن أمكن، هو تشكيل وتكوين المقياس باستخدام أكثر من بند واحد.

وكما هو مذكور في الفصل الثالث، فإن معامل «أوميغا» هو بديل لمعامل «ألفا» الذي قد يكون مناسباً عندما لا يتم الوفاء بالافتراضات الخاصة باختبارات معادلة التكافؤ الأساسي لـ «تاو» essentially tau-equivalent tests. وكما ذكرت أيضاً في ذلك الفصل، إن «دان وآخرين» (Dunn et al. (2014 و«بيترز» (Peters (2014 يقدمون إرشادات لاستخدام البيئة الإحصائية (R) لحساب معامل «أوميغا».

الخطوة الثامنة: تحقيق الحد الأمثل لطول المقياس:

تأثير طول المقياس على الثبات:

في هذه المرحلة من عملية تطوير المقياس، يكون لدى الباحث تجمع من البنود التي تظهر ثبات بدرجة مقبولة. ويتأثر معامل «ألفا» للمقياس بخاصيتين هما: مدى التغير بين البنود وعدد البنود في المقياس. بالنسبة للبنود التي لها ارتباطات بين المقياس والبند يساوي متوسط الارتباط بين البنود وبعضها البعض (أي، البنود التي تكون نموذجية إلى حد ما)، سيؤدي إضافة المزيد منها إلى

زيادة قيمة معامل «ألفا» وحذف المزيد منها سيقول قيمة معامل «ألفا». وبوجه عام، تكون المقاييس الأقصر (القصيرة إلى حد كبير) جيدة لأنها تلقي عبئاً أقل على المجيبين. ومن ناحية أخرى، فإن المقاييس الأطول تكون جيدة لأنها تميل إلى أن تكون أكثر ثباتاً. ومن الواضح أن تعظيم أحد هذه الأصول من شأنه أن يقلل من الآخر. ولذلك؛ ينبغي على مطور المقياس أن يعطي بعض التفكير في المفاضلة المثلى بين الإيجاز «الاختصار» والثبات.

إذا كان ثبات أي مقياس منخفضاً للغاية، فإن الإيجاز لا يكون فضيلة أو نقطة قوة. وقد يكون الأشخاص، في الواقع، أكثر استعداداً للإجابة على مقياس مكون من ثلاث بنود أكثر من استعدادهم للإجابة على مقياس مكون من عشرة بنود. ومع ذلك، إذا لم يتمكن الباحث من تعيين أي معنى للدرجات التي تم الحصول عليها من نسخة المقياس الأقصر، فلن يتم تحقيق أي شيء. وبناءً على ذلك، ينبغي أن تقتصر مسألة المقايضة بين الثبات والإيجاز على الحالات التي يكون فيها لدى الباحث «ثباتاً احتياطياً». وإذا كان الأمر كذلك، في الواقع، فقد يكون من المناسب تبني مقياس أقصر مقابل التضحية بقليل من الثبات.

#### آثار استبعاد وإسقاط البنود السيئة:

سواء كان إسقاط واستبعاد البنود «السيئة» في الواقع يزيد أو يخفض قليلاً قيمة معامل «ألفا» فإن ذلك يعتمد على مدى ضعف البنود التي سيتم إسقاطها واستبعادها وعلى عدد البنود الموجودة في المقياس. لننأمل هنا التأثير الذي يسببه عدد كبير أو قليل من البنود والتي تكون «جيدة» بنفس القدر (أي التي لها ارتباطات مماثلة مع نظيراتها): في حالة وجود عدد أقل من البنود، فإنه سينتج عن إضافة كل بند أو حذفه تغيير أكبر في معامل «ألفا». إذا كان متوسط الارتباط المتبادل بين البنود وبعضها البعض لعدد أربعة بنود هو (0.50)، فإن معامل «ألفا» سوف يساوي (0.80). وإذا كان هناك ثلاثة بنود فقط بمتوسط ارتباط متبادل بين بعضها البعض قدره (0.50)، فإن معامل «ألفا» سوف ينخفض إلى (0.75). وفي حالة وجود خمسة بنود لها نفس متوسط الارتباط فإن معامل «ألفا» لها سيبليغ (0.83). وبالنسبة للمقاييس التي تتألف من 9 و10 و11 بنوداً مع متوسط ارتباطات متبادلة بين البنود وبعضها البعض قدره (0.50)، فإن معامل «ألفا» سيبليغ (90.0، 91.0، 92.0) على التوالي. في الحالات الأخيرة، فإن معاملات «ألفا» ليست فقط أعلى ولكن أيضاً أقرب بكثير في القيمة من بعضها البعض.

إذا كان للبند ارتباط أقل من المتوسط بدرجة كافية مع البنود الأخرى، فإن إسقاطه واستبعاده سيرفع قيمة معامل «ألفا». وإذا كان متوسط ارتباطه بالبنود الأخرى أقل قليلاً (أو يساوي أو يزيد) من المتوسط الكلي، فإن الاحتفاظ بالبند سيزيد قيمة معامل «ألفا». وقد ذكرت أعلاه أن مقياس مكون من أربعة بنود سيحصل على معامل «ألفا» قدره (0.80)، مع وجود متوسط ارتباط متبادل بين البنود وبعضها البعض قدره (0.50). ولكن ما مقدار انخفاض متوسط ارتباط أحد البنود بالبنود الثلاثة الأخرى للمساعدة في اتخاذ قرار باستبعاد هذا العنصر بدلاً من إلحاق الضرر بمعامل «ألفا»؟ أولاً، فكر في متوسط الارتباط بين البنود الذي ينبغي أن يكون عليه لمقياس مكون من ثلاثة بنود لتحقيق معامل «ألفا» قدره (0.80). سيلزم أن يكون متوسط الارتباط بين البنود هو (0.57). لذلك، فبعد استبعاد أسوأ البنود الأربعة، ستحتاج البنود الثلاثة المتبقية إلى متوسط ارتباط متبادل فيما بينها قدره (0.57). حتى تحافظ على قيمة معامل «ألفا» تساوي (0.80). وسيكون لثلاثة بنود متوسط الارتباط فيما بينها أقل من (0.57) معامل «ألفا» أقل من أربعة بنود يبلغ متوسط الارتباط المتبادل بين بعضها البعض (0.50). وإذا افترضنا أن أفضل ثلاثة بنود في مقياس مكون من أربعة بنود كان متوسط الارتباط فيما بينها يساوي (0.57)، فإن متوسط الارتباط بين البند المتبقي (وبالتالي الأسوأ) والبنود الثلاثة الأخرى لابد أن يكون أقل من (0.43) لكي يؤدي استبعاده إلى زيادة قيمة معامل «ألفا» فعلياً. [إن وجود ثلاثة بنود يبلغ متوسط الارتباط المتبادل فيما بينها (0.57) وبند واحد يبلغ متوسط ارتباطه مع البنود الثلاثة الأخرى (0.43) ينتج عن ذلك متوسط ارتباط متبادل بين البنود الأربعة بشكل عام قدره (0.50)]. بالنسبة لأي قيمة أكبر من (0.43)، فإن وجود بند رابع يكون أكثر فائدة من خفض متوسط الارتباط بين البنود يؤدي إلى الضرر. وبالتالي، يجب أن يكون البند «السيئ» أسوأ قليلاً من البنود الثلاثة الأخرى (0.57 - 0.43 = 0.14) التي تستحق استبعادها.

والآن، فكر في الموقف عندما يكون هناك مقياس مكون من عشرة بنود وله معامل «ألفا» قدره (0.80). أولاً وقبل كل شيء، يجب أن يكون متوسط الارتباط بين البنود حوالي (0.29) فقط، مما يوضح الطريقة التي يعوض بها وجود المزيد من البنود الارتباطات الأضعف فيما بينها. ولكي يحقق مقياس مكون من تسعة بنود نفس قيمة معامل «ألفا»، يجب أن يكون متوسط الارتباط بين البنود حوالي (0.31). ولا بد أن يكون للبند «السيئ» متوسط ارتباط متبادل مع البنود التسعة المتبقية (0.20) أو أقل حتى يتسنى إدراجه كبند عاشر لسحب وخفض المتوسط الكلي للارتباط

المتبادل بين البنود إلى ما دون (0.29). قد يؤدي الفشل في خفض المتوسط إلى ما دون هذه القيمة إلى استفادة معامل "ألفا" من إدراج وتضمين البند. ويبلغ متوسط فرق الارتباط بين البنود التسعة "الجيدة" وبين البند الوحيد "السيئ" في هذه الحالة ( $11.0 = 20.0 - 31.0$ )، وهو فرق أصغر من فرق الموجود في مثال المقياس المكون من البنود الأربعة.

#### ضبط وإصلاح طول المقياس:

كيف يمكن للمرء أن يقوم بإصلاح أو «ترقيع» tinkering طول المقياس في الممارسة العملية؟ من الواضح أن البنود التي تسهم بأقل قدر في الاتساق الداخلي العام ينبغي أن تكون أول البنود التي يتم النظر في استبعادها. ويمكن تحديد هذه البنود بعدة طرق. يُظهر إجراء الثبات باستخدام برنامج "SPSS" وخيار "ألفا" "ALPHA" الخاص بالبروتوكول PROC CORR في برنامج "SAS" تأثير حذف كل بند على معامل "ألفا" الكلي. وعادةً ما يكون البند الذي يكون لإسقاطه وحذفه أقل تأثير سلبي أو أكبر تأثير إيجابي على معامل «ألفا» هو أفضل بند ينبغي إسقاطه وحذفه أولاً. ويمكن أيضاً استخدام ارتباطات المقياس بالبند كمقياس للبنود القابلة للاستنفاد (أي القابلة للحذف والإسقاط). ويجب أولاً إسقاط واستبعاد تلك البنود التي لديها أدنى مستوى من الارتباطات بين المقياس والبنود. كما يوفر البرنامج الإحصائي "SPSS" أيضاً الارتباط المتعدد التريبيعي squared multiple correlation لكل بند، يتم الحصول عليه عن طريق انحدار البند على جميع البنود المتبقية. وهذا يكون تقدير لتشاركية communality البند، ومدى تقاسمه التباين مع البنود الأخرى. وكما هو الحال مع ارتباطات المقياس بالبند، فإن البنود التي لها ارتباطات متعددة تريبيعية في أدنى مستوى ممكن لابد أن تكون هي البنود الرئيسية المرشحة للاستبعاد. وبصفة عامة، تتقارب وتتلاقى هذه المؤشرات المختلفة لجودة البند. وعادةً ما يصاحب الارتباط الضعيف للمقياس بالبند ارتباط متعدد تريبيعي منخفض وانخفاض صغير، أو حتى زيادة، في معامل "ألفا" عند إسقاط واستبعاد البند. يؤثر طول المقياس على دقة معامل "ألفا". ومن الناحية العملية، فإن معامل "ألفا" المحسوب هو تقدير للثبات يعتمد على مدى ملاءمة افتراضات القياس للبيانات الفعلية. وقد لوحظ بالفعل أن معامل "ألفا" يزداد عندما يتم تضمين وإدراج المزيد من البنود (ما لم تكن بنود فقيرة نسبياً). وبالإضافة إلى ذلك، يزداد ثبات «ألفا» كتقدير للثبات مع زيادة عدد البنود. وهذا يعني أن معامل "ألفا" المحسوب لمقياس أطول سيكون له فترة ثقة أضيق حوله أكثر من

معامل «ألفا» المحسوب لمقياس أقصر. وفي جميع التطبيقات، سيسفر المقياس الأطول عن قيم لمعامل «ألفا» أكثر تشابهاً مقارنةً بما سيسفر عنه مقياس أقصر. وينبغي أن تؤخذ هذه الحقيقة في الاعتبار عند تحديد طول أو إيجاز (اختصار) المقياس أثناء عملية التطوير.

وأخيراً، من المهم أن نتذكر أنه يجب تضمين هامش أمان في معامل «ألفا» عند محاولة تحسين طول المقياس. قد ينخفض معامل «ألفا» إلى حد ما عندما يتم (تطبيق) المقياس على عينة أخرى غير العينة المستخدمة في تطويره.

### تقسيم العينات:

إذا كانت عينة التطوير كبيرة بما فيه الكفاية، فقد يكون من الممكن تقسيمها إلى عيتين فرعيتين. ويمكن أن تكون إحداها بمثابة عينة التطوير الأساسية، ويمكن استخدام الأخرى للتحقق من النتائج المتحصل عليها. لذلك، على سبيل المثال، يمكن استخدام البيانات من العينة الفرعية الأولى لحساب معامل «ألفا»، وتقييم البنود، وضبط أو إصلاح طول المقياس، والتوصل إلى نسخة نهائية من المقياس تبدو مثالية. ويمكن بعد ذلك، استخدام العينة الفرعية الثانية ل تكرار هذه النتائج. ولن يستند اختيار البنود التي ينبغي الاحتفاظ بها على الإطلاق إلى العينة الفرعية الثانية. وبالتالي، فإن معاملات «ألفا» وغيرها من الإحصاءات المحسوبة لهذه المجموعة لن تظهر تأثيرات الصدفة chance، مثل تضخم معامل «ألفا»، الذي تمت مناقشته في وقت سابق. إذا ظلت معاملات «ألفا» ثابتة إلى حد ما عبر العيتين الفرعيتين، فيمكنك أن تكون أكثر ارتياحاً لافتراض أن هذه القيم غير مشوهة عن طريق الصدفة. وبطبيعة الحال، من المرجح أن تكون العيتين الفرعيتين أكثر تشابهاً من عيتين مختلفتين تماماً. ومن المرجح أن تكون العينات الفرعية، والتي يتم تقسيمها بشكل عشوائي من عينة تطوير بأكملها، ممثلة للمجتمع الإحصائي؛ وعلى النقيض من ذلك، قد تكون عينة جديدة تماماً ممثلة لمجتمع إحصائي مختلف بعض الشيء. كما أن فترات جمع البيانات للعيتين الفرعيتين لا يفصل بينها الزمن، في حين أن عينة التطوير وعينة منفصلة تماماً، على الدوام، يتم فصل الفترات الزمنية في جمع البيانات. وعلاوة على ذلك، فإن أي شروط خاصة قد يكون تم تطبيقها على جمع البيانات لعينة فرعية واحدة يتم تطبيقها بالمثل على العينة الفرعية الأخرى. ومن الأمثلة على الظروف الخاصة التعرض لأفراد بحث معينين، والإعدادات المادية، ووضوح طباعة الاستبيانات. كذلك، قد تكون العيتان الفرعيتان هما المجموعتان الوحيدتان اللذان يكملان الإجابة

على بنود المقياس مع كافة البنود من التجمع الأصلي للبنود التي تم رفضها في نهاية المطاف. وإذا كانت البنود المرفوضة قد مارست أي تأثير على الاستجابات على بنود المقياس، فإن هذه البنود ستكون قابلة للمقارنة بالنسبة لكلتا العيّنتين الفرعيتين.

وعلى الرغم من التشابه الفريد بين العينات الفرعية الناتجة، فإن تكرار النتائج عن طريق تقسيم عينة التطوير يوفر معلومات ذات قيمة حول استقرار المقياس. وتختلف العيّنتان الفرعيتان في جانب رئيسي واحد: ففي حالة العينة الفرعية الأولى التي استندت على بياناتها في اختيار البنود، توجد فرصة لخلط عوامل عدم الاستقرار والصدفة مع التغيرات الموثوق بين البنود. ولا توجد فرصة من هذا القبيل لإسناد أو عزو نتائج الصدفة بصورة منهجية «منتظمة» إلى الثبات بالنسبة للمجموعة الثانية لأن بياناتها لم تؤثر على اختيار البنود. وهذا الاختلاف المهم هو سبب كافٍ لتقييم المعلومات التي يمكن أن يوفرها تقسيم العينات في هذه المرحلة من تطوير المقياس.

إن الطريقة الأكثر وضوحاً لتقسيم عينة كبيرة بالقدر الكافي هو تقسيمها إلى نصفين. ومع ذلك، إذا كانت العينة صغيرة جداً بحيث لا يمكن أن يسفر تقسيمها عن نصفين كبيرين بشكل كافٍ، فإنه يمكنك تقسيمها بشكل غير متساوٍ. ويمكن استخدام العينة الفرعية الأكبر في العملية الأكثر أهمية لتقييم البنود وبناء المقياس واستخدام العينة الفرعية الأصغر للتحقق من صلاحية المقياس cross-validation.

#### تمارين:

افترض أنك تقوم بتطوير مقياس «الخوف من الثعابين» الذي يحتوي على صيغ استجابة «ليكرت» مكون من ستة خيارات، وتم تطبيقه على 300 شخص. على الرغم من أن المزيد من البنود سيكون أمراً مرغوباً فيه لتطوير المقياس الفعلي، إلا أنه يمكنك إكمال الخطوات التالية لهذه التمارين:

١- قم بإنشاء تجمع من عشرة بنود على شكل مقياس «ليكرت».

٢- بالنسبة لكل بند قمت بكتابته، قم بتقدير قيم مقياس «ليكرت» التي سيتم إقرارها من قبل «الشخص العادي» (أي، لا يوجد فوبيا من الثعابين أو ساحر الثعابين).



٣- اختر بنداً من تجمع البنود تعتقد أنه قد يستخلص استجابة متطرفة من شخص عادي، ثم أعد صياغتها للحصول على استجابة أكثر اعتدالاً.

٤- قم بتوليد وتكوين عشرة بنود أخرى من نوع مقياس «ليكرت» لقياس مفهوم بنائي آخر غير الخوف من الثعابين، ثم قم بمزج أو خلط هذه البنود عشوائياً مع البنود العشرة الأصلية واطلب من بعض الأصدقاء أن يشارروا إلى ما يعتقدون حول ما يهدف كل بند من البنود إلى قياسه.

٥- ما هي إيجابيات وسلبيات بناء البنود التي يكون لها عدد زوجي مقابل البنود التي يكون لها عدد فردي من خيارات الاستجابة؟

٦- باستخدام إما مفهوم الخوف من الثعابين أو المفهوم البنائي الأساسي الذي تقوم عليه المجموعة الثانية المكونة من عشرة بنود، قم بتدوين السلوكيات التي يمكن ملاحظتها بشكل مباشر والتي يمكن استخدامها للتحقق من صدق مقياس يقيس ذلك البناء وشرح كيف يمكنك استخدام البيانات السلوكية للتحقق من الصدق.

٧- ما قيمة معامل «ألفا» بالنسبة للمقياس إذا كان متوسط الارتباط المتبادل بين البنود العشرة لمقياس الخوف من الثعابين يساوي  $1(0,30)$ ؟

٨- كيف يمكنك استخدام العينات المقسمة لتقدير معامل «ألفا» للمقياس والتحقق من الصدق عبر العينات؟

ملاحظات:

$$١- الإجابة: ألفا = [١٠ * ٠,٣٠] / [٩ * ٠,٣٠ + ١] = ٠,٨١$$

## الفصل السادس

### التحليل العاملي

في الفصل الثاني، عند مناقشة النماذج النظرية المختلفة التي يمكن أن تصف علاقة بنود المقياس بالمتغير الكامن، تمت الإشارة إلى النموذج العاملي العام. هذا النموذج لم يفترض أن متغيراً كامناً واحداً فقط هو مصدر كل التغيرات covariation بين البنود. ولكن بدلاً من ذلك، يسمح هذا النموذج لمتغيرات كامنة متعددة لأن تكون كأسباب للتباين في مجموعة من البنود.

لتوضيح كيف يمكن لأكثر من متغير كامن أن يكون أساساً لمجموعة من البنود، سنعيد النظر في حالة محددة، وإن كان افتراضياً، تمت مناقشته في الفصول السابقة. يمكن تعريف العديد من المفاهيم البنائية constructs التي تهتم علماء الاجتماع والسلوك إجرائياً عند مستويات متعددة معينة. وتعتبر مفاهيم التكيف النفسي، والتأثير، والتأثير السلبي، والقلق، والقلق الاجتماعي، وقلق مواجهة الجمهور أمثلة على الظواهر الهرمية. حيث يمكن لكل مفهوم من هذه المفاهيم أن يستوعب تلك التي تتبعه في القائمة، مع إمكانية تطوير مقاييس عند كل مستوى معين. افتراضياً، يمكن للبنود ذات الصياغة المختلفة بأطر زمنية وخيارات استجابة مختلفة تتعلق إما بمستوى معين أو متوسط أو عام من هذا التتابع. ونأمل أن يقوم مطور المقياس باختيار صيغ البنود التي تتوافق مع المستوى المحدد للمتغير. عندئذ، يمكن استخدام التحليل العاملي لتقييم نجاح عملية الاختيار هذه.

لجعل هذا المثال أكثر تحديداً، خذ في الاعتبار مجموعة مكونة من 25 بنداً متعلقة بالتأثير. وتذكر أن أحد افتراضات نظرية الاستجابة للبند Item Response Theory هو أن البنود التي تشكل مقياساً هي أحادية البعد - أي أنها تتعلق بمفهوم بنائي واحد فقط. وبالتالي، فإن ما يهتما هو ما إذا كانت هذه البنود الـ 25 ينبغي أن تشكل مقياساً عاماً أو عدة مقاييس خاصة. هل جميع البنود الـ

25 تشكل مقياساً واحداً أم أنه من الأنسب أن يكون هناك مقاييس منفصلة لحالات التأثير المختلفة، مثل الاكتئاب والنشوة والعداء والقلق، وما إلى ذلك؟ بل ربما حتى سيكون من الأفضل تقسيم بنود التأثير الإيجابية والسلبية (على سبيل المثال، “سعيد” مقابل “حزين” للاكتئاب أو “متوتر” مقابل “الهدوء” للقلق) إلى مقاييس منفصلة. كيف نعرف ما هو الأنسب للبنود محل الدراسة؟ وبالأحرى، فإن السؤال الذي يطرح نفسه هو: هل مجموعة البنود المتعلقة بالعديد من حالات التأثير ناشئة من متغير كامن واحد أو من العديد من المتغيرات الكامنة؟

محاولة الإجابة على هذه الأسئلة مع استبعاد أسلوب التحليل العاملي والاعتماد فقط على الأساليب التي تمت مناقشتها في الفصول السابقة ستكون مهمة شاقة. يمكننا حساب معامل ألفا على كامل مجموعة بنود المزاجية، علماً أن معامل ألفا يخبرنا عن كمية التباين المشترك في مجموعة من البنود. إذا كانت ألفا منخفضة، فقد نبحت عن مجموعات فرعية من البنود التي ترتبط بقوة أكبر مع بعضها البعض. على سبيل المثال، قد نشك في أن بنود التأثير الإيجابية والسلبية لا ترتبط مع بعضها البعض وأن الجمع بينها يؤدي إلى خفض ألفا. لذا يفترض أن تكون قيم ألفا لهذه المجموعات الفرعية الأكثر تجانساً (كلها إيجابية أو سلبية التأثير) أعلى. قد نفترض بعد ذلك أنه حتى المجموعات الفرعية الأكثر تجانساً (على سبيل المثال، فصل القلق عن الاكتئاب بالإضافة إلى الإيجابي من السلبي) يجب أن تكون قيم ألفا لها أعلى. ومع ذلك، في مرحلة ما، قد نشعر بالقلق أيضاً من أن هذه المقاييس الأكثر تحديداً والمتجانسة سوف ترتبط ارتباطاً وثيقاً مع بعضها البعض لأنها كانت فقط مجرد جوانب مختلفة لنفس حالة التأثير، وهذا من شأنه أن يوحي بأن بنود تلك المقاييس تنتمي إلى نفس المقياس وليس إلى مقاييس منفصلة. يجب التأكيد على أن ألفا عالية نسبياً ليست ضماناً على أن جميع البنود تعكس تأثير متغير كامن واحد. فإذا كان هناك مقياس مكون من 25 بنداً، حيث 12 بنداً تعكس في المقام الأول متغيراً كامناً واحداً والـ 13 بنداً الأخرى تعكس متغيراً كامناً آخر، فإن مصفوفة الارتباط لجميع البنود يجب أن يكون لها بعض القيم العالية وبعض القيم المنخفضة. لذا يفترض أن تكون الارتباطات القائمة على بندين يمثلان نفس المتغير الكامن عالية، وتلك التي تعتمد على بنود تعكس بشكل أساسي متغيرات كامنة مختلفة ينبغي أن تكون منخفضة نسبياً. ومع ذلك، قد يكون متوسط الارتباط بين البنود مرتفعاً بما يكفي لإنتاج قيمة مقبولة لـ ألفا لمقياس مكون من 25 بنداً. في الواقع، للحصول على قيمة ألفا تساوي 80.0 لمقياس بهذا الطول، يجب أن يكون متوسط الارتباط بين البنود فقط 14.0. وبالتالي، مع وجود بنود تقيس نفس

المفهوم البنائي مرتبطة بمقدار 29.0 (ارتباط متوسط مقبول ولكنه ليس كبيراً بشكل غير معتاد لبنود داخل مجموعة فرعية متجانسة) ووجود بنود بين مجموعات مرتبطة بمقدار 00.0، لا تزال إمكانية الحصول على ألفا بمقدار 80.0 تقريباً قائمة. ولكن، بالطبع، يكون ألفا مناسباً فقط لمجموعة بنود أحادية البعد، لذلك قيمة الثبات reliability score (أي قيمة ألفا) هذه ستكون مضللة كمؤشر على الاتساق الداخلي أو أحادية البعد (الذي نعرفه لا ينطبق في هذا المثال الافتراضي). إن التحليل العاملي، موضوع هذا الفصل، هو أداة تحليلية مفيدة يمكن أن نخبرنا، بما لا يمكن لمعاملات الثبات أن تفعل، عن الخصائص المهمة للمقياس. يمكن أن يساعدنا تطبيقاً أو عملياً في تحديد عدد المفاهيم البنائية أو المتغيرات الكامنة أو العوامل وراء مجموعة بنود ما.

#### نظرة عامة حول التحليل العاملي:

يخدم التحليل العاملي عدة أغراض متصلة. وأحد وظائفه الرئيسية، كما لوحظ سابقاً، هو مساعدة الباحث في تحديد عدد المتغيرات الكامنة وراء مجموعة من البنود. وبالتالي، في حالة بنود التأثير البالغ عددها 25 بنداً، يمكن أن يساعد التحليل العاملي الباحث في تحديد ما إذا كانت هناك حاجة إلى بناء مفهوم بنائي واحد واسع أو عدة مفاهيم بنائية لتوصيف مجموعة البنود. أيضاً، يمكن أن يوفر التحليل العاملي وسيلة لتفسير التباين بين العديد من المتغيرات الأصلية المتصلة نسبياً (على سبيل المثال، 25 بنداً) باستخدام عدد قليل نسبياً من المتغيرات المنشأة (أي العوامل). وهذا يؤدي إلى تكثيف المعلومات بحيث يمكن حساب التباين باستخدام عدد متغيرات أقل. على سبيل المثال، بدلاً من الحاجة إلى 25 درجة (قيمة) لوصف كيفية إجابة المجيبين على البنود، قد يكون من الممكن حساب عدد أقل من الدرجات (ربما حتى واحدة) بناءً على دمج البنود. الغرض الثالث من التحليل العاملي هو تحديد المحتوى الموضوعي أو معنى العوامل (أي المتغيرات الكامنة) التي تفسر التباين بين مجموعة أكبر من البنود. يتم تحقيق ذلك عن طريق تحديد مجموعات من البنود تتغير covary مع بعضها البعض وتبدو أنها تُعرف بمتغيرات كامنة ذات معنى. إذا، على سبيل المثال، نشأ عاملان من تحليل بنود التأثير البالغ عددها 25 بنداً، فإن محتوى البنود الفردية التي تشكل مجموعات العوامل هذه يمكن أن تعطي فكرة عن المتغيرات الكامنة التي تمثلها العوامل. الوظيفة الرابعة للتحليل العاملي ترتبط بالوظائف الثلاثة المذكورة سابقاً. يساعد التحليل العاملي في تحديد البنود ذات الأداء الأفضل أو الأسوأ. وبالتالي، تلك البنود الفردية التي لا تنسجم مع أي من فئات

البنود المشتقة عاملياً أو التي تتناسب مع أكثر من واحدة من تلك الفئات العاملية يمكن تحديدها والنظر في إزالتها.

تقدم الأقسام التالية ملخصاً مفاهيمياً للتحليل العاملية، ويمكن للقراء الذين يريدون التعرف على المعالجة الحسابية للتحليل العاملية بشكل أعمق الرجوع إلى المراجع المخصصة لهذا الموضوع، مثل Cureton (1983) أو Gorsuch (1983) أو Harman (1976) أو McDonald (1984).

أمثلة على الطرق المماثلة للمفاهيم التحليلية للعوامل:

للحصول على تصور بديهي بما يقوم به التحليل العاملية، يمكننا النظر في عمليتين أقل رسمية ولكنهما متماثلتان تقريباً مع تلك التي ربما تكون أكثر دراية بها. أول هذه العمليات يتم استخدامها في بعض الأحيان في إدارة الموارد البشرية لتحديد الموضوعات المشتركة بين القضايا المحددة التي تبدو متنوعة والتي قد تهم أعضاء الفريق أو زملاء العمل.

مثال رقم 1:

افترض أن شركة صغيرة جديدة تريد تحديد الخصائص التي يعتقد موظفوها أنها مهمة لزملائهم في العمل للتحلي بها. إنهم يعتقدون أن تحديد ومكافأة الخصائص ذات القيمة على نطاق واسع سوف تلعب دوراً هاماً في تهيئة بيئة عمل متناغمة وتعاونية. وتقوم الشركة باستقطاب أخصائي موارد بشرية لمساعدتهم. يجمع هذا الأخصائي، الذي سنطلق عليه اسم جيم Jim، موظفي الشركة العشرة معاً ويوضح أنه يود منهم أن يفكروا في خصائص زملائهم الموظفين الذين يعتبرونها مهمة في مجموعة التفاعلات التي قد تكون لديهم في العمل، بدءاً من وضع المقترحات والتقارير معاً إلى التفاعل مع العملاء المحتملين معاً إلى مشاركة طاولة في الكافتيريا - المجموعة الكاملة من التفاعلات التي قد تكون لدى الموظفين. يقترح جيم أنه لبدء العملية، يكتب كل موظف على قصاصات ورق منفصلة أكبر عدد من الخصائص المهمة التي يمكنهم تحديدها.

بعد عدة دقائق، يكتب خلالها الموظفون أفكارهم، يطلب «جيم» من موظف متطوع قراءة إحدى أفكاره للمجموعة. تقول أليس Alice إن إحدى الخصائص التي كتبتها كانت «مستعد لمشاركة الأفكار». شكرها جيم وطلب منها أن تقوم بتعليق الورقة التي بها تلك الفكرة على الحائط.

موظف آخر، يدعى بيل Bill، يقرأ أحد خصائصه: “روح الدعابة” وهي، أيضاً، مسجلة على الحائط. تستمر هذه العملية مع قيام كل موظف على حدة بذكر كل الخصائص التي قام بتدوينها. وبهذه الطريقة، يحدد الموظفون بمفردهم مجموعة متنوعة من الخصائص التي يعتبرونها مهمة بالنسبة لزملاء العمل. بعد القيام بذلك، يقومون بلصق قصاصات الورق التي تحتوي على تلك الخصائص على الحائط. من بين تلك الخصائص المذكورة ما يلي:

لديه النية لمشاركة الأفكار.	ودي.
روح الدعابة.	يمكن الاعتماد عليه.
دائماً لديه الأدوات المناسبة للوظيفة.	يهتم بالتفاصيل.
ذكي.	لديه ذاكرة حديدية.
متقن.	متحفظ.
مُجد.	يعرف الكثير من العملاء المحتملين.
يفكر بمنطق.	جاد.
تجده في المواقف الصعبة.	لديه شخصية.
يستعد للمهام.	متعلم بشكل جيد.
يترك انطباعاً جيداً لدى العملاء.	جدير بالثقة.
لا يحاول الاستحواذ على كل الثناء.	مهتم بمظهره.
يعمل بروح الفريق.	يجيد الأسلوب القصصي.
مرح.	حاد الذكاء.
لديه سيارة رائعة.	شخص لديه عقيدة.
لديه خبرة كبيرة في هذا النوع من العمل.	لديه الاستعداد لساعات طويلة إذا استدعى الأمر ذلك.

وتستمر هذه العملية لبعض الوقت، وسرعان ما تتم تغطية الجدار بأكثر من 30 ملصقة ورقية، كل منها تذكر سمة كان يعتقد الموظف أنها مهمة. ثم بعد ذلك، يسأل جيم المشاركون أي الخصائص يرون أنه من الممكن ضمها مع بعضها البعض. تشير كثيرين إلى أن «ذكي» و«حاد الذكاء» متشابهان. يأخذ جيم الورقة التي كتبت عليها كلمة «ذكي» وينقلها إلى جانب كلمة «حاد الذكاء». ويقترح فرانك أن الخاصية «متعلم بشكل جيد» يجب أن تكون أيضاً جزءاً من هذه المجموعة. وتتم إضافة العديد من الخصائص الأخرى إلى نفس مجموعة العبارات. ثم تلاحظ كارلا أن «ودي» و«يترك انطباعاً جيداً لدى العملاء» متشابهة مع بعضها البعض وتختلف عن المجموعة الأخرى من العبارات التي تم تشكيلها مسبقاً. إنها تقترح أن يتم تجميع هاتين الخاصيتين في مجموعة جديدة، ثم يضاف «مرح» أيضاً إلى هذه المجموعة الثانية. الخاصيتان «متحفظ» و«مهتم بمظهره» تشكلان نواة المجموعة الثالثة إلا أن إحدى الموظفين تعتقد أن «متحفظ» سيكون من الأفضل ضمها إلى «يستعد للمهام» بدلاً من «مهتم بمظهره». واستمرت هذه العملية حتى قام جيم والموظفون بتشكيل عدة مجموعات من الخصائص. تقريباً، تم وضع كل خاصية موصوفة في مجموعة ما.

ثم يطلب جيم من المشاركين إعطاء أسماء -كلمة أو عبارة وصفية قصيرة— لكل مجموعة من العبارات. مجموعات العبارات المختلفة هذه أعطيت مسميات مثل «الذكاء»، «المظهر»، «الوعي»، «الشخصية»، «الاعتمادية»، وغيرها من المسميات الأخرى. افتراضياً، كل مجموعة من العبارات تمثل مفهوماً رئيسياً متعلقاً بتصورات الموظفين عن خصائص بعضهم البعض.

## مثال رقم 2:

بعد عدة سنوات، قررت الشركة تكرار التمرين. المدراء لديهم شك في أن الأمور قد تغيرت بما فيه الكفاية بحيث لا تظل كل الفئات المحددة سلفاً ذات صلة. جيم، ميسر الموارد البشرية، غير متواجد. لذا قررت كارول، أحد المديرين التنفيذيين للشركة، أن الطريقة الأسهل للحصول على معلومات مماثلة قد تتمثل في تطوير استبيان يحتوي على بيانات مثل تلك التي توصل إليها الأشخاص في التمرين السابق. سيطلب من الموظفين الإشارة إلى مدى أهمية كل خاصية، باستخدام العبارات «لا على الإطلاق»، و«إلى حد ما» و«جداً» كخيارات استجابة لهم. طبقت هذه الاستبيانات على الموظفين، الذين بلغ عددهم الآن ما يقرب من 150 موظفاً. وعندما جمعت كارول

الاستبانات بعد تعبئتها، اطلعت عليها لمعرفة الخصائص الأكثر أهمية. وكان أحد الأشياء التي لاحظتها هو أن هناك خصائص مختلفة مهمة لموظفين مختلفين، ولكن هناك بعض الخصائص المحددة صنف بشكل مشابه مع بعضها البعض. على سبيل المثال، الأشخاص الذين اعتقدوا أن الخاصية “يهتم بالتفاصيل” كانت مهمة، كان من المحتمل أن يعتقدوا أن الخاصية “يستعد للمهام” مهمة أيضاً، والأشخاص الذين لا يعتبرون أحد هذه العناصر مهمة عادةً لا يعتبرون الأخرى مهمة أيضاً. بحثت كارول في سبل للاستفادة أكثر من كومة الاستبانات وفكرت في كيفية فهمها. تذكرت كيف أنه خلال التمرين الأصلي الذي أجري قبل عدة سنوات باستخدام قصاصات الورق على الحائط، ظهر أن هناك مجموعات أكثر مما هو مطلوب بالفعل. اعتقدت كارول أن بعض العبارات الفردية لا قيمة لها إلى حد ما، وأحياناً إذا كان لدى الشخص نفسه أكثر من واحدة من هذه العبارات التي لا قيمة لها، فإن فئة كاملة لا قيمة لها ستتج. وتساءلت عما إذا كانت هناك طريقة لتحديد عدد الفئات التي سيستغرقها تنقيح معظم ما يعتقد الموظفون بشأن زملائهم في العمل. وكتمرين، حاولت البحث عن مجموعات أخرى من العناصر (الخصائص)، مثل تلك الاثنين اللتين لاحظتهما سابقاً، والتي تميل لأن تكون متماثلة عند الموظفين. بشكل رئيسي، بحثت كارول عن مجموعات من العناصر المتشابهة لا تستند فقط إلى محتواها ولكن إلى مدى تقييم الموظفين لها أيضاً. استغرق هذا وقتاً كبيراً، ولم تكن كارول متأكدة بالفعل من أنها حصلت على كل مجموعات العبارات المهمة، لكنها شعرت أنها قادرة على جمع بعض الأفكار المثيرة للاهتمام بهذه الطريقة من الاستبانات.

أوجه القصور في هذه الطرق:

كلا هذين المثالين يشبهان من الناحية النظرية التحليل العاملي ولكن مع بعض الاختلافات الهامة. في كلتا الحالتين، النتيجة هي إعادة تنظيم كمية كبيرة من المعلومات المحددة في مجموعة أكثر قابلية للإدارة من فئات أكثر عمومية ولكن ذات معنى. افتراضياً، كل من عمليات إعادة التصنيف في فئات قد نتج عنها عدد قليل من الأفكار التي اكتسبت الكثير مما غطته العبارات الفردية العديدة. ومع ذلك، كان لكلا النهجين أوجه قصور واضحة إلى حد ما. في المثال الأول، كانت هناك سيطرة بسيطة على جودة البيانات الناتجة. هذا القصور ينبع من حقيقة أن الأشخاص الذين يطورون البنود يختلفون في مستوى مهاراتهم في تكوين الأفكار ذات الصلة في عبارات موجزة. بعض الأشخاص منفتحون أكثر من غيرهم وقد ينتهي بهم المطاف إلى توليد المزيد من العبارات؛ ومع ذلك، فليس الحال دائماً أن يكون أعضاء المجموعة الأكثر انفتاحاً هم الأكثر بصيرة. ولهذه



الأسباب وغيرها، تؤدي هذه العملية غالبًا إلى عبارات غامضة أو غير ذات صلة أو سخيفة. ولذا قد يكون من الصعب إزالة البنود ضعيفة الصياغة أو تحسينها. اعتمادًا على مرونة المجموعة المشاركة في التمرين، قد يكون من غير المناسب رفض هذه العبارات دون مجادلة مؤلفيها. ونتيجة لذلك، يتم منحهم مصداقية أكبر بقدر العبارات الأفضل التي تعكس المزيد من الأفكار المهمة. على الرغم من التفاوت في الجودة من بند إلى آخر، إلا أنه قد يتم التعامل مع الجميع بشكل متساوٍ يزيد أو ينقص.

إذا تم تقديم العديد من البنود السخيفة ولكنها متشابهة، فمن المحتمل أن تشكل فئة تستند فقط إلى تشابهها. ربما يتم ترتيب الفئات حسب الأولوية، ولكن يتم ذلك عادةً بتوافق الآراء، واعتمادًا على من قام بإعداد العبارات، قد يكون هناك إجماع عن وصف بعض الفئات بأنها تافهة. علاوة على ذلك، تشير تجربتي مع هذا النوع من التمارين إلى وجود ميل قوي لوضع كل عبارة في فئة معينة. يبدو أن وجود العديد من الفئات الواضحة ومن ثم وجود عبارة واحدة أو اثنتين يتامى لا تنتمي لأي من الفئات يجعل الأشخاص لا يوقفون عملية التصنيف حتى يتم توزيع تلك العبارات على الفئات، حتى وإن كان التوزيع في لحظتها غير ملائم. علاوة على ذلك، حتى عندما يمكن لشخص ما تحديد أي عبارات محددة تجسد فئة ما، فليس من الواضح بالضرورة تحديد أيها أفضل أو أسوأ.

لذلك؛ كيف كان أداء هذا الأسلوب بشكل عام بالنسبة إلى وظائف التحليل العملي المحددة سابقاً؟ من الواضح أن هذه الطريقة جعلت من الصعب تحديد البنود ذات الأداء الأفضل أو الأسوأ. تتعارض مراقبة جودة البنود مع الطبيعة الجماعية لهذه المهمة وقد تخاطر بالإساءة إلى بعض المشاركين. وعلى الرغم من أن الطريقة الموصوفة للتو لها بعض الفائدة في تحديد عدد المتغيرات الكامنة التي تستند إليها العبارات التي قام الموظفون بإنشائها، مع إسناد هذا التحديد إلى الانطباعات الذاتية للموظفين، كما هو الحال، يبدو أقل من المثالي. على الرغم من أن الموظفين قد يكون لديهم رؤية قيمة في مكان عملهم، فمن غير المرجح أن يكون لديهم خبرة في تنظيم هذه الأفكار إلى مجموعات متماسكة تمثل مفاهيم بنائية constructs مهمة. علاوة على ذلك، هناك القليل من الأساس الموضوعي لتحديد ما إذا كان الموظفون كمجموعة قد قاموا بعمل جيد أو سيء لتحديد عدد المتغيرات الكامنة وراء العبارات المنشأة، لأن العملية تعتمد فقط على معايير ذاتية. وفيما يتعلق بتجميع المعلومات، تظهر مشكلة مماثلة. على الرغم من أنه قد يتم تجاهل بعض البنود (أو حتى

الفئات بأكملها)، إلا أن المعايير المستخدمة هي بالضرورة ذاتية ولا توجد سوى وسائل قليلة للحكم بشكل موضوعي على ما إذا كانت الخيارات الجيدة قد اتخذت. الاعتبار الوحيد الذي قد لا تؤديه هذه الطريقة بشكل سيء للغاية هو تعريف المحتوى الموضوعي أو معنى العوامل. إذا كانت العوامل ذات مصداقية (وهي مفتوحة للسؤال)، فإن فحص محتوى العبارات المقدمة قد يعطي نظرة ثاقبة في المفاهيم البنائية الأساسية التي كان الموظفون يدركون أهميتها.

المثال الثاني تجنب بعض أوجه القصور هذه. يمكن لكارول التخلص من البنود التي لا ترى مناسبة لكونها غير ذات صلة، على الرغم من أن هذا يضيف عبئاً كبيراً على حكمها. على الأقل كانت عملية تأييد البنود أكثر ديمقراطية إلى حد ما. على كل شخص أن يقيم كل بند دون التعرض لخطر استبعاد زميل العمل. وبالتالي، فإن عملية تحديد البنود ذات الأداء الأفضل أو الأسوأ ربما كانت على الأقل أكثر اتساقاً. علاوة على ذلك، لم يتم تحديد المجموعات من خلال مجرد شعور بالتشابه الواضح في العبارات ولكن عن طريق الأدلة على أن الناس تفاعلوا مع البنود التي تم تصنيفها في مجموعات متشابهة بطريقة مشتركة. أي إن التشابه كان سمة البنود (مجموعات معينة من التي تبدو أنها تثير تصورات مماثلة)، وليس المستجيبون (الذين تباينوا في استجاباتهم على أي بند محدد). رؤية بند واحد في مجموعة على أنه غير مهم يعني أنه من المحتمل بشكل كبير لنفس الفرد رؤية البنود الأخرى في نفس المجموعة على أنها غير مهمة. قد يرى موظف آخر أن هذه البنود نفسها مهمة بشكل متنسق. المسألة الحاسمة هي أنه، بغض النظر عن تقييم الفرد للأهمية، هناك ميل لأن تكون متسقة خلال العبارات داخل المجموعة. في الواقع، كان هذا هو الأساس الذي شكلت عليه كارول المجموعات. وبالتالي، ربما قد تم تحسين عملية تحديد عدد المتغيرات الكامنة إلى حد ما من خلال اعتماد هذه المنهجية المنقحة القائمة على الاستبيان. بطبيعة الحال، فإن تقييم البنود عن طريق الفحص البصري لـ 150 استبياناً سيكون أمراً شاقاً إلى حد كبير، ومن المحتمل أن نظام التصنيف لكارول لم يكن أكثر الطرق الممكنة كفاءة. ومن الأسئلة التي تثار: ما مدى الاتساق المطلوب للبنود لاعتبارها مجموعة؟ كم عدد حالات قيام موظف واحد بإعطاء تقييمات متباينة (أي اتفاق واختلاف في الأهمية) لبندين في نفس المجموعة المحتملة يمكن أن تقبلها كارول؟ من الصعب القول إلى أي مدى تكون هذه الطريقة قد حققت الوظائف المتبقية - تركيز المعلومات وتحديد المحتوى الموضوعي أو معنى العوامل. يعتمد كلاهما على مدى جودة العمل الذي كانت كارول

قادرة على القيام به لإيجاد واستبعاد البنود الضعيفة، والتحديد بشكل صحيح الموضوعات المهمة، ومن ثم استخدام البنود داخل المجموعات لتفسير المفاهيم البنائية المحتملة.

باختصار، من الواضح أن هاتين الطريقتين تتركزان الكثير مما هو مرغوب فيه لو كان الهدف هو تحقيق الوظائف المذكورة أعلاه والتي ترتبط عادةً بالتحليل العملي. وللاّنصاف، قد لا يكون الهدف من المنهجيات الموصوفة (خاصة الأولى) تحقيق تلك الوظائف، ولكن يبدو أن بعض الأهداف المماثلة لما ينجزه التحليل العملي تحفز القيام بممارسات مثل التمرين الأول الذي تم وصفه سابقاً. في حين أنه تم القبول بهذه الحجة، فإن تمرين الموظف بالمقارنة مع التحليل العملي يخدم الأغراض: (أ) تقديم توضيح أكثر سهولة لما يمكن أن يحققه التحليل العملي و(ب) تسليط الضوء على أوجه القصور المحتملة لنهج ذاتي وغير رسمي لتحديد المتغيرات الأساسية.

#### الوصف المفاهيمي للتحليل العملي:

التحليل العملي عبارة عن فئة من الإجراءات التي تنجز نفس النوع من التصنيف مثل الأساليب المذكورة أعلاه ولكنها تفعل ذلك وفقاً لمجموعة من العمليات الأكثر تنظيماً وتوفر معلومات أكثر وضوحاً يمكن لمحلل البيانات استخدامها في إصدار الأحكام. وكالطرق الموصوفة آنفاً، يحدد التحليل العملي فئات من العبارات المتماثلة. تتمثل المهمة الأولى للمحلل في تحديد عدد الفئات التي تكفي لالتقاط الجزء الأكبر من المعلومات الموجودة في مجموعة العبارات الأصلية.

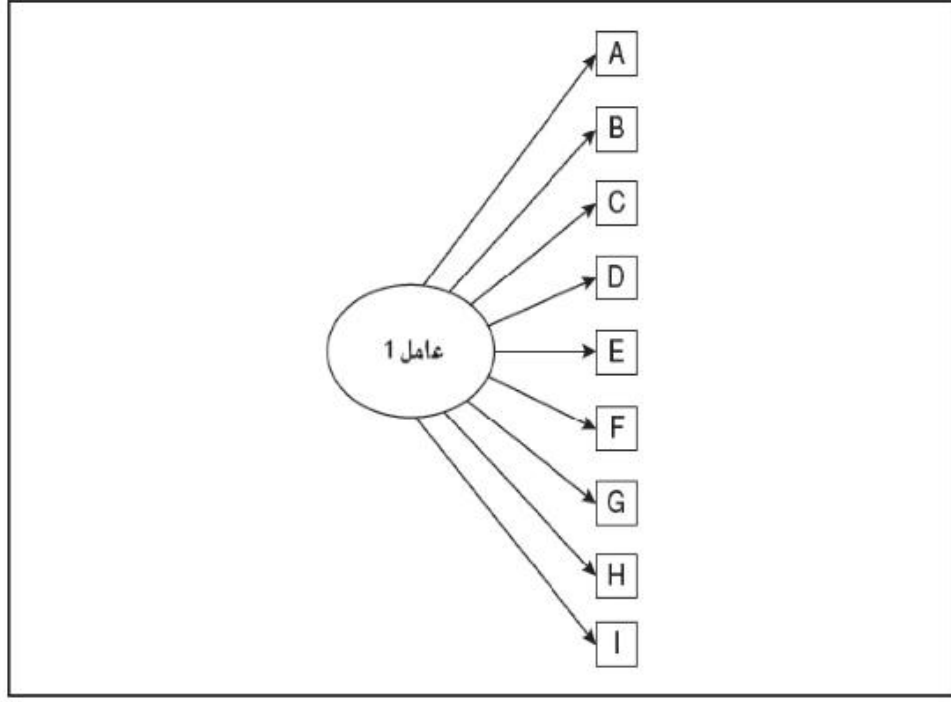
#### استخلاص العوامل:

يمكن للشخص أن ينظر إلى التحليل العملي كبداية على افتراض أن فئة واحدة كبيرة تحتوي على جميع البنود هو كل المطلوب (بمعنى، أن مفهوم أو فئة واحدة تكفي للأخذ في الاعتبار نمط الاستجابات). عندئذٍ يقوم بتقييم مقدار الارتباط بين البنود الفردية التي يمكن أن يفسرها ذلك المفهوم الوحيد. يقوم التحليل بعد ذلك بإجراء فحص لمعرفة مدى نجاح فرضية المفهوم الوحيد. إذا ظهر أن مفهومًا أو فئة واحدة لم تقم بالعمل اللازم لحساب التباين covariation، بين البنود فإن التحليل العملي يرفض الفرضية المبدئية. ثم يحدد المفهوم الثاني (أي المتغير الكامن أو العامل) الذي يفسر بعض التباين المتبقي بين البنود. تستمر هذه العملية إلى أن يكون مقدار التباين الذي لم تفسره مجموعة العوامل صغيراً.

## العامل الأول:

كيف يتم ذلك؟ تبدأ العملية بمصفوفة الارتباط لجميع البنود الفردية. باستخدام هذه المصفوفة كنقطة بداية، يقوم التحليل العامل بفحص أنماط التباين الممثلة بالارتباطات بين البنود. ما يلي هو وصف مفاهيمي، وتم حذف بعض التفاصيل الرياضية لمصلحة الوضوح، لذلك لا ينبغي أن يؤخذ ذلك حرفيًا حيث إنه فقط مجموعة العمليات التي تقوم عليها تحليلات العوامل المولدة بالحاسوب.

كما ذكرنا سابقاً، تتضمن العملية الفرضية المبدئية للمفهوم الفردي الذي يمكن وبشكل كافٍ أن يأخذ بالاعتبار نمط الارتباطات بين البنود. وهذا بمثابة تأكيد مؤقت على أن النموذج الذي يحتوي على متغير كامن واحد (أي عامل واحد)، مع وجود مسار منفصل ينبثق منه إلى كل بند من البنود، هو تمثيل دقيق للعلاقات السببية (انظر الشكل رقم ٦-١). هذا يعني كذلك أن مثل هذا النموذج يمكن أن يأخذ في الاعتبار الارتباطات بين البنود. لاختبار هذا الافتراض من الناحية المفاهيمية، يجب أن يحدد برنامج التحليل العامل العلاقة بين كل بند والعامل الذي يمثل المتغير الكامن الفردي، ومن ثم النظر إلى ما إذا كان يمكن إعادة إنشاء الارتباطات المرصودة بين البنود عن طريق ضرب المسارات التي تربط كل زوج من المتغيرات عبر العامل. ولكن كيف يمكن للبرنامج حساب الارتباطات بين استجابات البنود المرصودة والعامل الذي يمثل متغير كامن لم يتم ملاحظته أو قياسه مباشرة؟

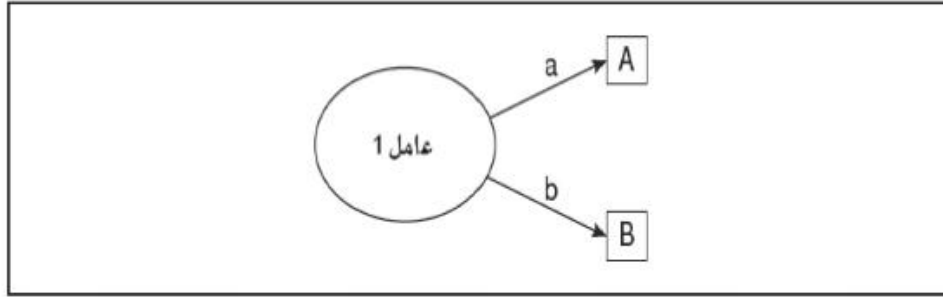


شكل (٦-١): نموذج العامل الواحد

أحد الأساليب هو افتراض أن مجموع جميع الاستجابات على البنود هو تقدير عددي معقول للمتغير الكامن الشامل الذي يُفترض أن يأخذ في الاعتبار الارتباطات بين البنود. وبشكل رئيسي، هذا المجموع الكلي هو تقدير «لدرجة» المتغير الكامن. ولأن الدرجات الفعلية لجميع البنود يُفترض في الوقت الحالي تحديدها من خلال متغير كامن واحد، فإن كمية المعلومات من جميع البنود (أي المجموع الكلي للمعلومات) هو تقدير معقول للقيمة العددية للمتغير الكامن. من السهل إلى حدٍ ما إضافة درجات البنود معاً إلى مجموع الدرجات وحساب ارتباطات البند - المجموع item-total correlations لكل بند منفصل مع إجمالي كل البنود. تعمل ارتباطات البند - المجموع كممثل للارتباطات بين البنود المرصودة والمتغير الكامن الغير مرصود (أي المسارات السببية من المتغير الكامن إلى البنود). مع القيم المحددة لهذه المسارات السببية، يمكن للشخص بعد ذلك حساب الارتباط بين البنود المتوقعة بناءً على نموذج العامل الواحد. تمثل هذه الارتباطات النموذجية إسقاطات لما يجب أن تكون عليه الارتباطات الفعلية بين البنود إذا كانت فرضية المتغير الأساسي الوحيد صحيحة. يمكن تقييم شرعية الفرضية من خلال مقارنة الارتباطات المتوقعة مع

الارتباطات الفعلية. هذا يعني طرح كل ارتباط متوقع من الارتباط الفعلي المقابل بناءً على البيانات الأصلية. يشير الفرق الكبير بين الارتباطات الفعلية والمتوقعة إلى أن نموذج العامل الواحد غير مناسب؛ ولا يزال هناك بعض التباين بين البنود لم يتم أخذه بالاعتبار.

قد يساعد النظر في هذا التسلسل لزوج واحد من البنود،  $A$  و  $B$ ، التي تعد جزءاً من مجموعة أكبر. أولاً، ستتم إضافة المجموعة الكاملة من البنود، بما في ذلك  $A$  و  $B$ ، معاً للحصول على درجة ملخصة، ثم سيتم حساب الارتباطات بين  $A$  والدرجة الإجمالية وبين  $B$  والدرجة الإجمالية. يُفترض أن هذين الارتباطين بين البند والمجموع الكلي يمثلان الارتباطات بين  $A$  و  $B$  والعامل، الذي يتوافق مع المتغير الكامن الأساسي. إذا كانت فرضية متغير كامن وحيد صحيحة، فسيكون لرسم تخطيطي للمسار يشتمل على  $A$  و  $B$  والعامل مسارات (a و b في الشكل 6-2) من العامل إلى كل من البندين  $A$  و  $B$ . ستكون قيم هذه المسارات هي الارتباطات بين البند والمجموع الكلي للبنود الموصوفة للتو. بناءً على مخطط المسار هذا، يجب أن يكون الارتباط بين  $A$  و  $B$  هي حاصل ضرب هذين المسارين. حساب هذا الارتباط المقترح بين  $A$  و  $B$  يتضمن الضرب البسيط.



شكل (٦-٢): نموذج مبسط لعامل واحد يتضمن بندين فقط

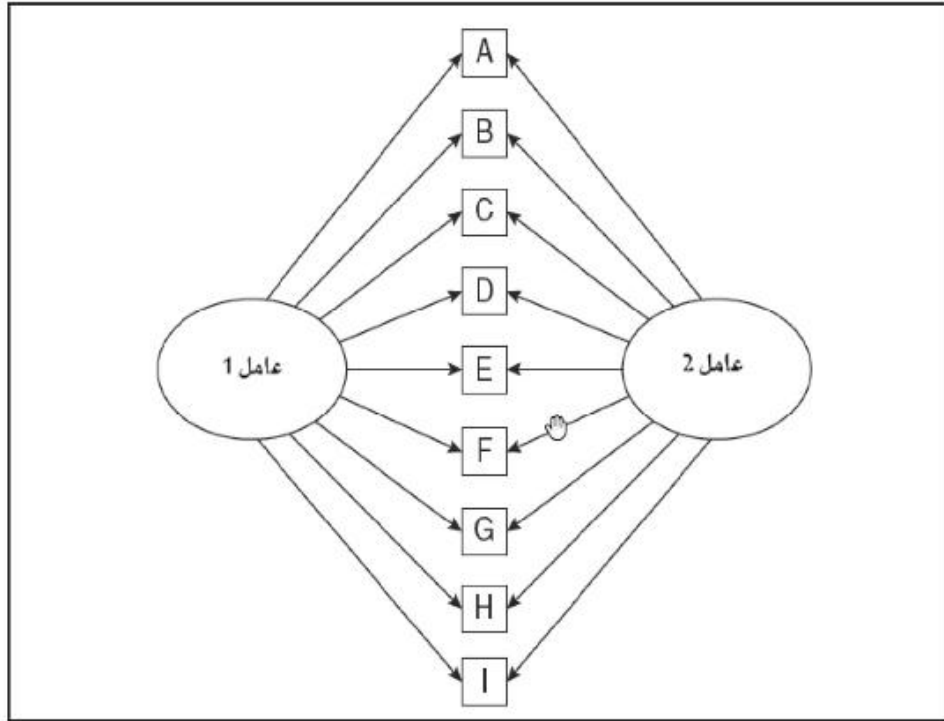
بمجرد حسابه، يمكن مقارنة الارتباط المقترح مع الارتباط الفعلي بين  $A$  و  $B$ . يمكن طرح الارتباط المقترح من الارتباط الفعلي للحصول على الارتباط المتبقي. قد يشير الارتباط المتبقي الكبير إلى أن الادعاء بأن متغير كامن وحيد هو السبب الوحيد للتباين بين  $A$  و  $B$  غير كافٍ.

العمليات التي تتم على مصفوفة الارتباط بأكملها في وقت واحد تقوم بالإجراء المذكور سابقاً لكل زوج ممكن من البنود. وبدلاً من الانتهاء إلى ارتباط متبقي واحد، يتم حساب مصفوفة كاملة من الارتباطات المتبقية (تسمى: مصفوفة البقايا residual matrix)، بحيث يمثل كل ارتباط متبقي

مقدار التغيرات بين زوج معين من البنود الموجودة فوق وما بعد التغيرات الذي يمكن لمتغير كامن واحد أن يفسره.

العوامل اللاحقة:

أصبح من الممكن الآن العمل على مصفوفة البقايا هذه بنفس الطريقة التي عولجت بها مصفوفة الارتباط الأصلية، واستخراج عامل ثانٍ مقابل لمتغير كامن جديد. مرةً أخرى، يمكن حساب الارتباطات بين البنود والمتغير الكامن الثاني (العامل ٢ في الشكل رقم ٦-٣)، وبناءً على تلك الارتباطات، يمكن إنشاء مصفوفة الارتباطات المقترحة.



شكل (٦-٣): نموذج بعاملين

تمثل هذه الارتباطات المقترحة مدى الارتباط الذي يجب أن يبقى بين البنود بعد أخذ العامل الثاني في الاعتبار. إذا استحوذ العامل الثاني على كل التغيرات المتبقي بعد استخراج العامل الأول، فينبغي أن تكون هذه القيم المتوقعة مماثلة للقيم التي كانت موجودة في مصفوفة البقايا المذكورة أعلاه. إذا لم يكن الأمر كذلك، فقد تكون هناك حاجة إلى مزيد من العوامل لحساب التغيرات المتبقي الذي لم يُنسب بعد إلى أحد العوامل.

يمكن أن تستمر هذه العملية، مع كل عامل جديد مستخرج من مصفوفة البقايا التي نتجت عن التكرار السابق، حتى يتم تحقيق مصفوفة تحتوي فقط على علاقات متبادلة صغيرة مقبولة. عند هذه النقطة، يمكن للشخص أن يتخذ قراراً أنه قد تم حساب جميع التغيرات المهمة بشكل أساسي وأنه لا توجد عوامل أخرى مطلوبة. من الممكن متابعة العملية حتى يتم الحصول على مصفوفة بقايا تتكون بالكامل من الأصفار. وهذا سيحدث عندما يكون عدد العوامل المستخرجة يساوي عدد البنود في التحليل العامل. وبمعنى آخر، ستنتم دائماً مجموعة من العوامل  $k$  من تفسير كل التغيرات بين مجموعة من البنود  $k$ .

تحديد عدد العوامل التي يجب استخلاصها:

يمكن أن يكون تحديد عدد العوامل التي يمكن استخراجها مشكلة معقدة (على سبيل المثال، Zwick & Velicer, 1986). وبطبيعة الحال، فإن الدافع الرئيسي لإجراء التحليل العامل هو الانتقال من مجموعة كبيرة من المتغيرات (البنود) إلى مجموعة أصغر (العوامل) تقوم بعمل معقول لالتقاط المعلومات الأصلية (أي تركيز المعلومات). ويمكن تحديد "العمل المعقول" بعدة طرق.

بعض الأساليب التحليلية للعوامل تستخدم معياراً إحصائياً، مثل تلك القائمة على التقدير بأسلوب الاحتمال الأعظم والتحليل العامل التوكيدي (والذي ستم مناقشته لاحقاً) القائم على طريقة نمذجة المعادلات البنائية. في هذا السياق، يشير مصطلح المعيار الإحصائي إلى حقيقة أن الأساليب الاستدلالية تُستخدم لتحديد ما إذا كان احتمال وجود نتيجة معينة ضئيلاً بدرجة كافية لاستبعاد حدوثها بالصدفة. هذا يعني إجراء اختبار لمعرفة ما إذا كانت البقايا المتبقية تحتوي على مقدار من التغيرات تكون إحصائياً أكبر من الصفر وذلك بعد استخراج كل عامل لاحق. إذا كان هذا هو الحال،



تستمر العملية حتى لم يعد الأمر كذلك. إن الاعتماد على معيار إحصائي بدلاً من الحكم الشخصي هو سمة جذابة لهذه الأساليب. ومع ذلك، في تطوير المقاييس، قد لا يتوافق هذا مع الهدف المطروح، والذي يتمثل في تحديد مجموعة صغيرة من العوامل التي يمكن أن تفسر التغيرات المهم بين البنود. تسعى الأساليب الإحصائية إلى الحصول على تحديد شامل للعوامل الكامنة وراء مجموعة من البنود. إذا كان هناك بعض مصادر التغيرات لم يتم حسابها من قبل أي من العوامل المستخرجة بعد، فإن مثل هذه الأساليب تستمر في استخراج المزيد من العوامل. وما يسعى إليه مطور المقياس في كثير من الأحيان بعد ذلك هو الاختزال في عدد العوامل. أي إنه في سياق تطوير المقياس، غالبًا ما نريد معرفة مصادر الاختلاف القليلة والأكثر نفوذًا التي تقوم عليها مجموعة من البنود، وليس كل مصدر ممكن يمكننا اكتشافه. عند تطوير المقياس، يقوم المرء عادةً بإنشاء قائمة طويلة من البنود أكثر من المتوقع للوصول إلى الأداة النهائية. والبنود التي لا تساهم في العوامل الرئيسية المحددة قد يتم حذفها في نهاية المطاف. هدفنا هو تحديد عدد قليل نسبيًا من البنود التي ترتبط بقوة بعدد صغير من المتغيرات الكامنة. وعلى الرغم من أن محلل البيانات الماهر يمكنه تحقيق هذا الهدف من خلال الأساليب التحليلية للعوامل باستخدام معيار إحصائي، إلا أن الباحث الأقل خبرة، ومن المفارقات، قد يكون أفضل من خلال استخدام إرشادات أخرى أكثر ذاتية ولكن من المحتمل أن تكون أقل غموضاً.

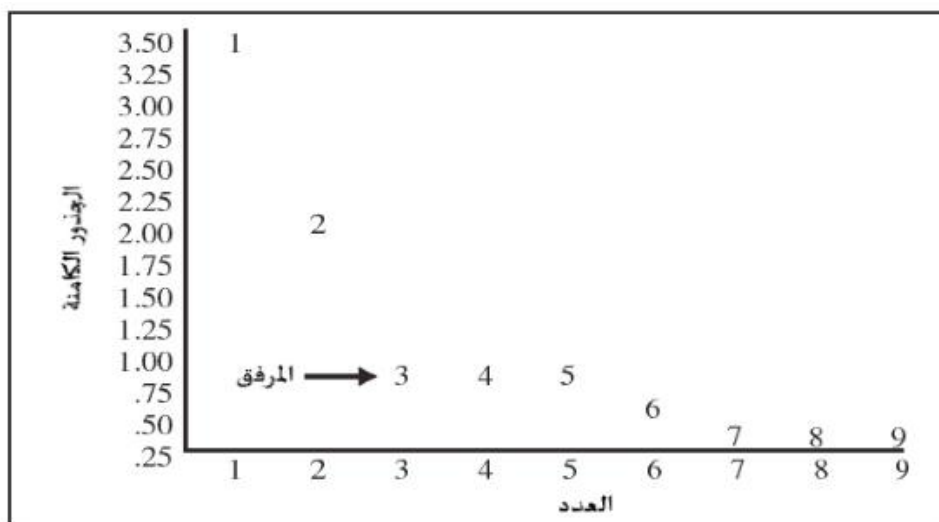
هذه الإرشادات الذاتية تعتمد نسبيًا على نسبة التباين الكلي بين البنود الأصلية التي يمكن لمجموعة من العوامل شرحها. وهذا هو الأساس نفسه الذي تستخدمه الأساليب الإحصائية. بينما في حالة المعايير غير الإحصائية (بمعنى، لا تستند إلى الاحتمالية)، يقوم محلل البيانات بتقييم مقدار المعلومات التي يحتوي عليها كل عامل تالٍ ويتخذ القرار حيال ذلك متى ما تم الوصول إلى نقطة تناقص النتائج. يشبه هذا تقريبًا تفسير الارتباطات (مثل معاملات الثبات) على أساس حجمها، معيار ذاتي، وليس على أساس قيمتها الاحتمالية  $p$ ، كمعيار إحصائي. اثنان من المبادئ التوجيهية غير الإحصائية المستخدمة على نطاق واسع للحكم عند استخراج عوامل كافية هي قاعدة الجذور الكامنة (Kaiser, 1960) eigenvalue واختبار الحصة (Cattell, 1966) scree test

يمثل الجذر الكامن (القيمة الكامنة) eigenvalue مقدار المعلومات التي تم التقاطها بواسطة العامل. بالنسبة لأنواع معينة من طرق تحليل العوامل (مثل تحليل المكونات الرئيسية principal

component analysis، والتي ستنم مناقشته لاحقاً في هذا الفصل)، فإن إجمالي كمية المعلومات في مجموعة من البنود يساوي عدد البنود. وبالتالي، في تحليل 25 بنداً، سيكون هناك 25 وحدة من المعلومات. الجذر الكامن لكل عامل يتوافق مع جزء من هذه الوحدات. على سبيل المثال، في حالة تحليل مكون من 25 بنداً، فإن العامل ذا الجذر الكامن 0.5 يمثل 20% (5/20) من إجمالي المعلومات؛ جذر كامن آخر بمقدار 5.2 تمثل 10% من إجمالي المعلومات، وهلم جرأً. نتيجة لهذه العلاقة بين كيفية قياس المعلومات وعدد البنود في التحليل هي أن الجذر الكامن بقيمة 0.1 يقابل  $1/k$  من التباين الكلي بين مجموعة من البنود. وبعبارة أخرى، فإن العامل (بافتراض تحليل المكونات الرئيسية) الذي يحقق جذر كامن تساوي 0.1، يحتوي على نفس النسبة من إجمالي المعلومات كما هو الحال في البند الفردي العادي. وبالتالي، إذا كان الهدف من التحليل العملي هو الوصول إلى عدد أقل من المتغيرات التي تلتقط بشكل كبير المعلومات الواردة في مجموعة المتغيرات الأصلية، فإن العوامل يجب أن تكون محملة بالمعلومات أكثر من البنود الأصلية. وفقاً لذلك، تؤكد قاعدة الجذر الكامن (Kaiser, 1960) أنه لا ينبغي الاحتفاظ بالعوامل ذات الجذور الكامنة التي أقل من 0.1 (وبالتالي، تحتوي على معلومات أقل من البند المتوسط). على الرغم من أن الأساس المنطقي لاستبعاد هذه العوامل أمر منطقي، فماذا عن العوامل التي تزيد قليلاً عن 0.1؟ هل العامل الذي يفسر معلومات أكثر بنسبة 1% من البند العادي حقاً يقدم نوعاً من المعلومات المركزة التي نلحقها؟ في كثير من الأحيان، تكون الإجابة “لا”، مما يوحي بأن قاعدة الجذور الكامنة قد تكون سخية للغاية كأساس للاحتفاظ بالعوامل. أعتقد أن هذا هو الحال عموماً في تطوير المقياس استناداً إلى الأساليب الكلاسيكية.

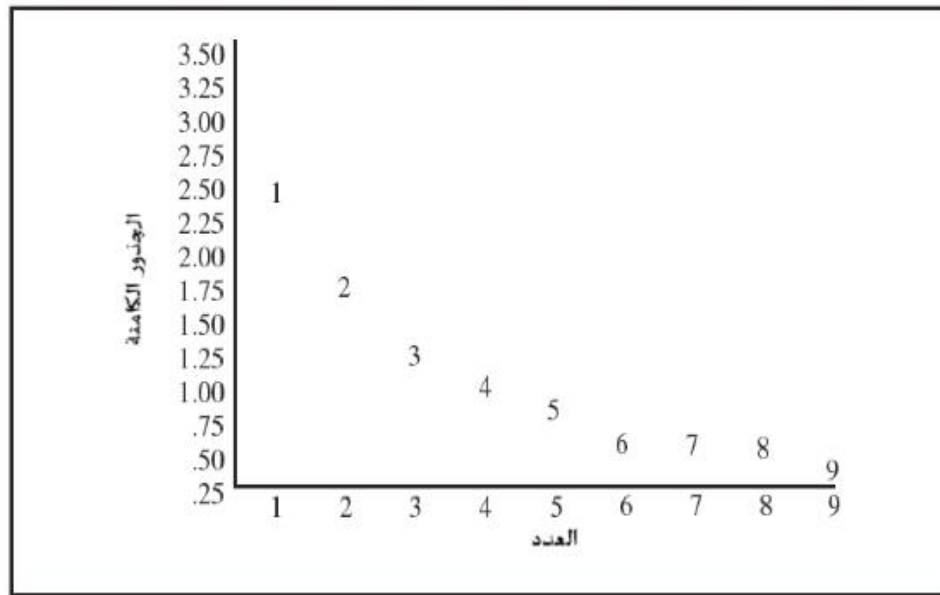
ويستند اختبار الحصاة (Cattell, 1966) scree test أيضاً على الجذور الكامنة ولكن يستخدم قيمها النسبية بدلاً من قيمها المطلقة كمعيار. ويستند إلى الرسم البياني للجذور الكامنة المرتبطة بالعوامل المتعاقبة. ونظراً لأن كل عامل، بعد العامل الأول، يتم استخلاصه من مصفوفة تمثل البقايا residuals الناتجة من استخراج العامل السابق (كما هو موضح سابقاً)، فإن كمية المعلومات في كل عامل تالٍ تكون أقل من سابقتها. اقترح كاتيل Cattell أنه يمكن تحديد العدد “الصحيح” من العوامل من خلال النظر في انخفاض كمية المعلومات (وبالتالي، في حجم الجذر الكامن) عبر العوامل المتعاقبة. عند رسم هذه المعلومات، سيكون لها شكل يتسم بجزء عمودي في الغالب على اليسار (يمثل الجذور الكامنة الكبيرة) ينتقل إلى جزء أفقي نسبياً على اليمين (مطابق

للجذور الكامنة الصغيرة). واعتبر كاتيل العوامل المرتبطة بالجانب الأيمن التي تمثل الجزء الأفقي من الرسم البياني مستهلكة لا حاجة لها. في المصطلحات العامة، الحصاة تصف الركام الذي يتجمع على الأرض بعد الانهيار الأرضي. يعني هذا المصطلح، إذن، أن الجزء الرأسي من الرسم البياني هو المكان الذي توجد فيه العوامل الجوهرية بينما الجزء الأفقي هو الحصاة، أو الركام، التي يجب التخلص منها. من الناحية المثالية، سيكون لتقدم العوامل نقطة تسقط عندها المعلومات فجأة، مع انتقال مفاجئ من رأسي إلى أفقي و"مرفق elbow" واضح (انظر الشكل ٦-٤). يدعو معيار Cattell إلى الإبقاء على تلك العوامل الكامنة فوق المرفق من الرسم البياني. في بعض الأحيان، لا يكون الانتقال مفاجئاً لكنه تدريجي (انظر الشكل رقم ٦-٥)، مع وجود منحنى خفيف يتكون من عدة عوامل تقع بين المناطق الرأسية والأفقية للرسم البياني. في مثل هذه الحالات، قد يكون تطبيق اختبار الحصاة لـ Cattell أمراً خادعاً ويتطلب اعتماداً أكبر على معايير ذاتية، مثل قابلية التفسير للعامل. يعتبر العامل قابلاً للتفسير إلى الحد الذي تبدو فيه البنود المرتبطة به متشابهة مع بعضها البعض وتجعلها ذات معنى من الناحية النظرية ومنطقية كمؤشرات لمفهوم بنائي متماسك.



شكل (٦-٤): رسم الحصاة scree plot يتضمن مرفق elbow واضح

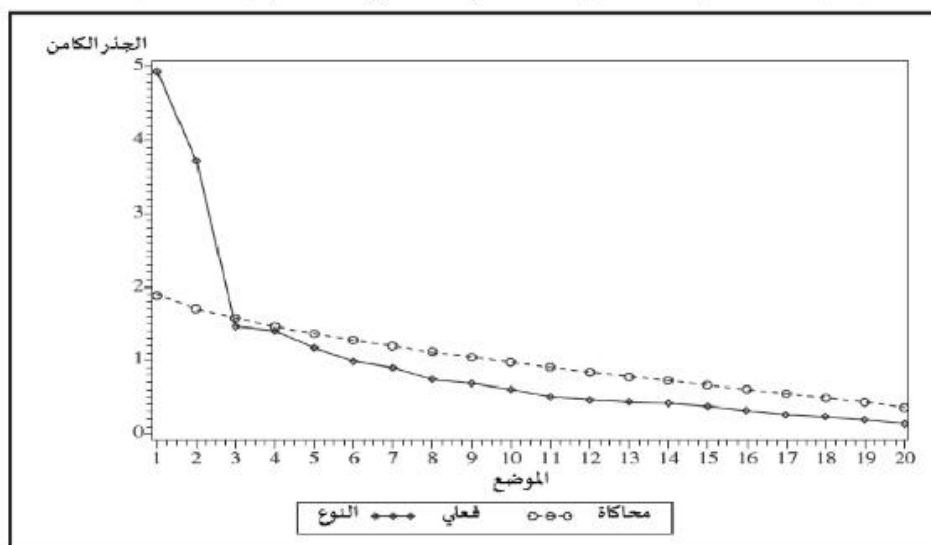
يوجد معيار إحصائي آخر، له شعبية متزايدة، يعتمد على التحليل الموازي analysis parallel (على سبيل المثال، Hayton, Allen, & Scarpello, 2004). المنطق خلف هذا النهج هو أن حجم الجذر الكامن للعامل الأخير المُحتَفَظ به يجب أن يتجاوز الجذر الكامن الذي تم الحصول عليه من بيانات عشوائية في ظل ظروف مماثلة. بمعنى آخر، في دراسة حقيقية تتضمن التحليل العاملي، قل مثلاً، 20 متغيراً مقيساً على 500 شخص، يجب أن يكون الجذر الكامن لأي عامل مُحتَفَظ به أكبر من الجذر الكامن المقابل الذي تم الحصول عليه من بيانات تم إنشاؤها عشوائياً مُرتَّبَةً لتمثيل 20 متغيراً تم قياسها على 500 شخص. تقوم برمجيات التحليل الموازي، والتي تتوفر على الإنترنت كوحدات ماكرو مطورة بواسطة المستخدم لمجموعة متنوعة من الحزم الإحصائية، بإنشاء عدد كبير من التوزيعات تتوافق مع حجم مجموعة البيانات الفعلية.



شكل (٥-٦): رسم الحصاة scree plot لا يتضمن مرفق elbow واضح

وبالتالي، في هذا المثال، سوف تولد الخوارزمية الكثير (يمكن للمستخدم عادةً تحديد العدد؛ 1000 غالباً العدد الافتراضي) من مجموعات البيانات العشوائية المكونة من 20 متغيراً و500 شخص. يتم استخراج أو استخلاص الجذور الكامنة لكل مجموعة من مجموعات البيانات المولدة، ويتم إنشاء توزيع الجذور الكامنة ضمن البرنامج للعامل الأول والعامل الثاني، وهكذا تستمر العملية لبقية العوامل. لكل من هذه التوزيعات، يتم تحديد قيمة ممثلة (على سبيل المثال، الوسيط). يتم

عرض المخرجات الرسومية، كما هو مبين في الشكل رقم (٦-٦). يحتوي هذا الشكل على خطين يربطان النقاط المتعلقة بالجذور الكامنة للعوامل المتتالية. الخط المستقيم تقريباً المتدرج تنازلياً من اليسار إلى اليمين يمثل وسيط الجذور الكامنة (عبر مجموعات البيانات العديدة المولدة بالحاسوب) للبيانات العشوائية التي لا ترتبط المتغيرات فيها بالعوامل الأساسية السائدة. الخط الآخر في الشكل هو رسم الحصة الفعلي، استناداً إلى البيانات الحقيقية محل الاهتمام. ويشار إلى عدد العوامل التي يجب الاحتفاظ بها من خلال تقاطع خطي البيانات الفعلية والعشوائية (المولدة). حجم الجذور الكامنة التي تقع فوق الخط الذي يمثل البيانات العشوائية أكبر من المتوقع بالصدفة بمفردها (أي إنها ذات دلالة إحصائية). وتلك القيم التي تقع أسفل خط البيانات العشوائية ليست معنوية (على سبيل المثال، يمكن توقع حدوثهم بالصدفة البحتة).



شكل (٦-٦): شكل بياني من التحليل الموازي يوضح حجم الجذور الكامنة للعوامل المتتالية من البيانات المولدة (المحاكاة) والبيانات الفعلية

في المثال الموضح في الشكل المصاحب، حقق العاملان الأولان من البيانات الفعلية قيمًا ذاتية أكبر من تلك للعاملين المقابلين استناداً إلى البيانات المولدة بشكل عشوائي. وبالتالي، فإن هذين العاملين الأولين فقط يفيان بمعايير الاستبقاء. لاحظ أن البيانات العشوائية المولدة تسفر عن عدد من العوامل بجذور كامنة أكبر من 0.1، مما يدل على عدم كفاية قاعدة الجذور الكامنة. يُظهر فحص خط البيانات الحقيقية أنه سيتم أيضاً الاحتفاظ بعاملين استناداً إلى معيار اختبار الحصة الذي يبقي

على العوامل الواقعة فوق نقطة “المرفق” (التي تتوافق مع العامل الثالث) من الرسم البياني. ولأن وحدات الماكرو (برمجيات حاسوبية) لإجراء التحليل الموازي في مختلف الحزم الإحصائية (بما في ذلك SAS وSPSS) جاهزة ومتاحة عبر الإنترنت دون قيود على استخدامها ودون مقابل، فإن تطبيق هذه الطريقة كوسيلة لاختيار عدد العوامل التي يجب الاحتفاظ بها أصبح ممارسة شائعة. ومن خلال تجربتي، فإن التوجيه الذي تقدمه هذه الطريقة يتماشى مع الأحكام التي يتم إصدارها بناءً على معايير ذاتية مثل رسم الحصة. ومع ذلك، فإن لها ميزة كونها تركز على معيار إحصائي، وبالتالي فهي تتمتع بقبول أوسع من الأساليب الذاتية. وفقاً لذلك، أوصي باستخدام التحليل الموازي كأساس جوهري لتحديد عدد البنود التي يجب استخراجها.

تجدر الإشارة إلى أنه تم تطوير التحليل الموازي لأول مرة للاستخدام مع المكونات الرئيسية بدلاً من العوامل المشتركة. وستتم مقارنة هذه الأساليب التحليلية المختلفة للعوامل لاحقاً. حالياً، نلاحظ فقط أنه بينما يتم حساب المكونات الرئيسية من مصفوفة الارتباط الأصلية، يتم حساب العوامل المشتركة من مصفوفة ارتباط مخفضة matrix correlation reduced تحتوي على تقديرات تم ضبطها لإزالة تباين الخطأ من المتغيرات. تم توسيع التحليل الموازي لاستخدامه مع العوامل المشتركة، ولكن كما لاحظ Timmermann و Seva-Lorenzo (2011)، فإن عدد الجذور الكامنة للبيانات الحقيقية التي تتجاوز تلك الخاصة بالبيانات العشوائية يميل إلى أن يكون أكبر في حالة العوامل المشتركة منها في حالة المكونات الرئيسية. المنطق وراء التحليل الموازي تم تطويره مع وضع المكونات في الاعتبار. وبالتالي، في بعض الحالات، فإن عدد العوامل التي يسفر عنها التحليل الموازي التي ينبغي الاحتفاظ بها من العوامل المشتركة قد يكون تقديراً مبالغاً فيه للأبعاد الحقيقية لمجموعة البيانات.

قد تظهر هناك مواقف عندما تدعم اعتبارات جوهريّة (مثل قابلية تفسير العوامل الناتجة) الاحتفاظ بالمزيد أو أقل من العوامل باستخدام طريقة مثل التحليل المتوازي. عندما يحدث هذا، نقترح تقديم نتائج الطريقة المستخدمة أولاً (على سبيل المثال، التحليل الموازي) مع الأساس المنطقي لاتخاذ القرار للابتعاد عن التوجيه الذي توفره هذه الطريقة. وإذا كانت المسائل الجوهرية مقنعة، فقد تكون هناك بعض الصعوبة في إقناع القراء والمراجعين بأنه قد تم التوصل إلى قرار

مناسب. من ناحية أخرى، إذا كانت الحجج التي تم تنظيمها دعماً لتجاهل النتائج الأولية ليست مقنعة، فقد لا يكون هناك ما يبرر رفض هذه النتائج.

#### تدوير العوامل:

الغرض من استخلاص العوامل هو بكل بساطة تحديد العدد المناسب من العوامل التي يجب فحصها. إن وضع المعلومات في التكوين أو الشكل الأكثر فهماً ليس هو الهدف من استخلاص العوامل. العوامل الخام الغير مدارة unrotated هي تجريدات رياضية لا معنى لها. وكقياس تقريبي، تخيل أنه قد طُلب منك وصف طول جميع الأشخاص في الغرفة. وقررت أن تفعل ذلك عن طريق اختيار شخص بشكل اعتباطي، حيث تم اختيار جو بشكل عشوائي، وقياس طوله، ووصف أي شخص آخر بأنه أطول أو أقصر من الشخص المرجعي، جو، بعدة بوصات. لذلك؛ قد يكون أحد الأشخاص أطول من جو بثلاث بوصات وآخر أقصر منه بمقدار بوصتين. في مثل هذه الحالة، جميع المعلومات حول الأطوال متوفرة بعدة طرق تقديمية للبيانات ولكن لم يتم تنظيمها بالطريقة الأكثر إفادة. سيكون من الأسهل على الناس تفسير البيانات إذا تم تحويلها إلى نموذج يمكن فهمه بسهولة، مثل طول كل فرد في الغرفة، معبراً عنه بالأقدام والبوصات. يشبه دوران العوامل التحول من حيث إنه يقدم البيانات المتاحة مسبقاً بطريقة أسهل للفهم.

قبل محاولة تفسير العوامل - للتأكد من المفاهيم البنائية أو المتغيرات الكامنة المقابلة للعوامل، على أساس البنود التي تم تحديدها مع كل عامل - عادةً ما يكون من الضروري إجراء تدوير للعوامل. يزيد دوران العوامل من قابلية التفسير من خلال تحديد مجموعات المتغيرات التي يمكن وصفها في الغالب بمتغير كامن واحد (أي، البنود المتشابهة من حيث إنها جميعاً لها علاقة قوية مع بعضها البعض، وبالتالي يتم تحديدها إلى حد كبير، بعامل واحد فقط). إن تدوير العوامل وقابلية التفسير الأفضل الناتجة عنه لا تتم بتغيير البنود أو العلاقات فيما بينها، ولكن عن طريق اختيار وجهات النظر التي يتم وصف البنود بناءً عليها.

من الصعب فهم فكرة أن نفس مجموعة المعلومات يمكن أن تكون أكثر أو أقل وضوحاً بناءً على وجهة النظر التي يتم فحص تلك المجموعة من خلالها. في الأقسام التالية، سيتم تقديم عدة أمثلة تهدف إلى توضيح هذه الفكرة.

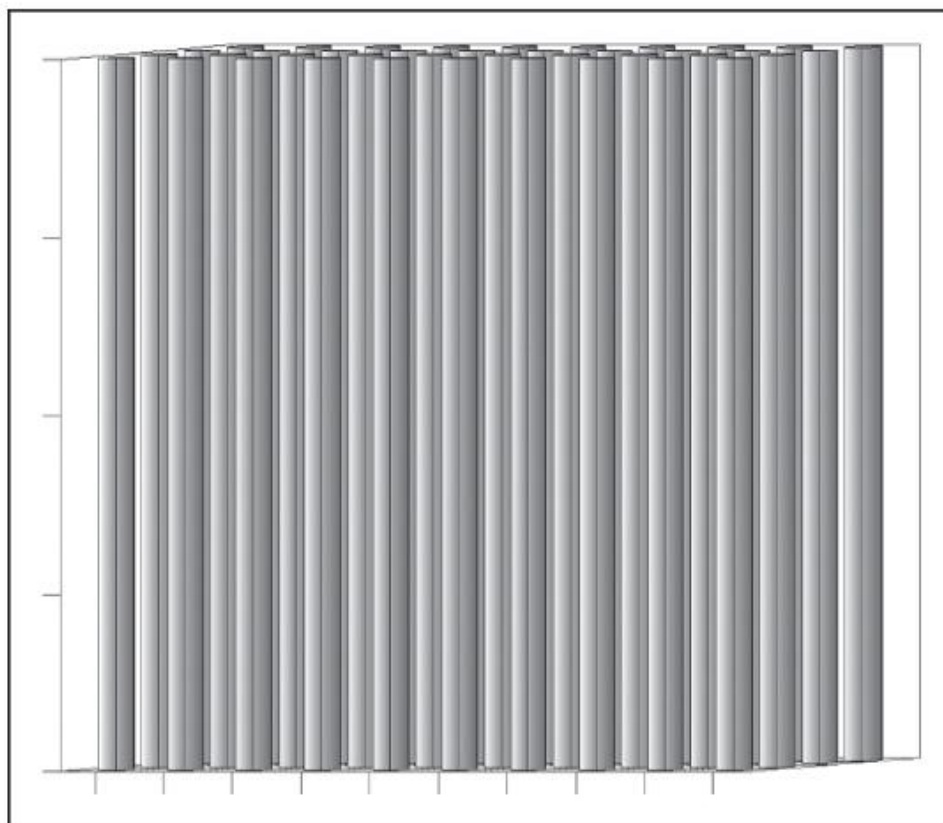
تتشابه أنماط العلاقات المتبادلة بين مجموعة من البنود مع المواقع المادية في الفضاء. كلما كان بندان أكثر ارتباطاً، كانت العلامتان الممثلتان لهذين البندين أقرب لبعضهما البعض. إذا قمنا بذلك للعديد من البنود، فإن المواقع الفعلية لعلاماتها ستتخذ نمطاً يمثل أنماط الارتباط بين المتغيرات (من الأسهل تصور ذلك إذا قصرنا أنفسنا على بعدين). إن تخيل مواقع الأجسام المادية التي تحددها القواعد الأساسية هو، بالتالي، طريقة أخرى للتفكير في البنود التي يتم تحديد ارتباطاتها بواسطة المتغيرات السببية الأساسية.

### التشبيه الأول لتدوير العوامل:

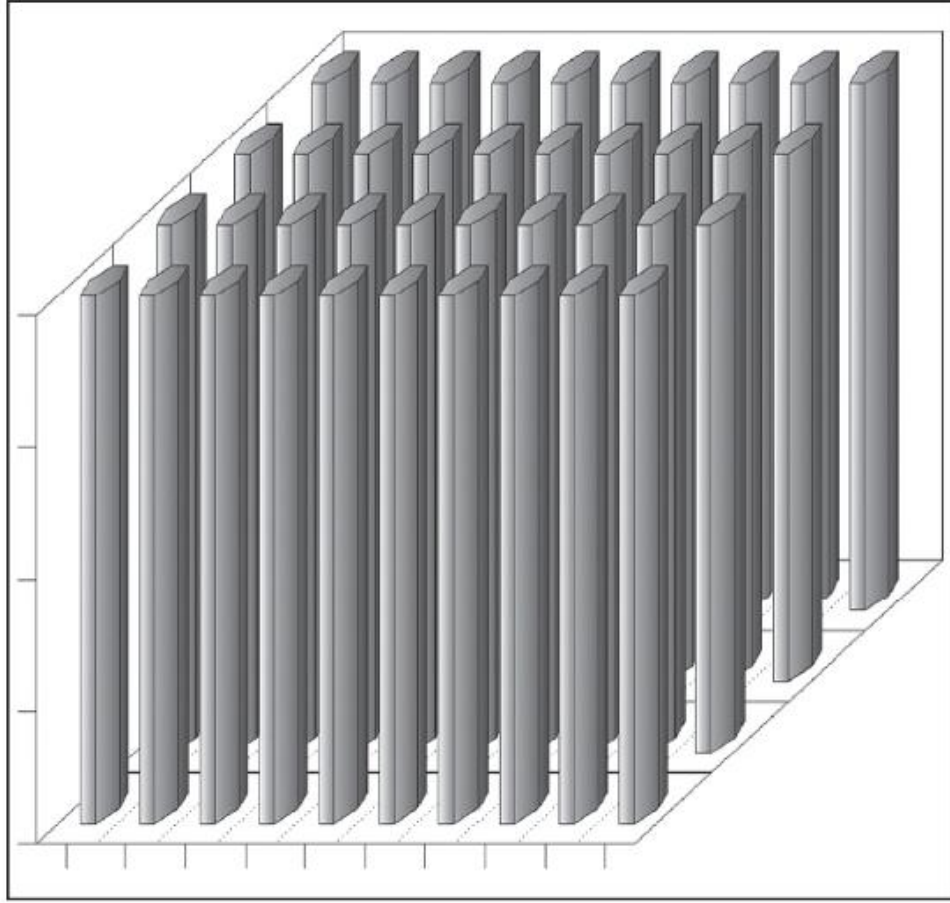
كيف يتيح لنا الدوران رؤية نمط بين المتغيرات الذي كان دائماً هناك ولكن لم يكن واضحاً؟ كقياس، تخيل مجموعة من الأشياء المنظمة جيداً، مثل مجموعة من الأعمدة مرتبة في سلسلة من الصفوف المتوازية والمنظمة. من الممكن الوقوف في مواقع معينة واختيار زوايا رؤية معينة بحيث يكون ترتيب الأشياء في الصفوف والأعمدة مخفياً بشكل كامل. ومع ذلك، تغيير زاوية الرؤية يمكن أن تكشف عن الترتيب الأساسي. فكر في كيفية ظهور ترتيب أربعة صفوف من 10 ركائز pillars عند عرضها من وجهات نظر مختلفة. إذا لم يكشف خط الرؤية عن أي من المحاور الطبيعية للترتيبات الخطية، فقد تظهر الأعمدة بشكل عشوائي. المشاهدة من موقع أعلى، أو التحرك بضعة أقدام يميناً أو يساراً، أو مجرد تحوّل نظر المرء أفقيّاً إلى خط الرؤية يمكن أن يؤدي إلى الكشف عن انتظام الركائز وإظهار ترتيبها، ويوضح الشكلان رقم (٦-٧) ورقم (٦-٨) هذه النقطة. هذه الأشكال تم إنشاؤها عن طريق رسم مواقع للأعمدة، وحرفياً، تغيير موقع الرؤية الذي من خلاله تم عرض الأعمدة. وبالتالي، فإن كلا الشكلين يمثلان نفس تكوين الأعمدة ويختلفان فقط في كيفية مشاهدتها. في الشكل الأول من هذه الأشكال، من الصعب في أحسن الأحوال تمييز أي تنظيم لترتيب الركائز. يقع خط الرؤية الذي يمر خلال المسافات بين الركائز على ركائز أخرى أبعد، جاعلاً التجمع يظهر ككتلة غير منظمة من الأشياء. يمثل الشكل الثاني من هذه الأشكال وجهة نظر بديلة. كل ركيزة، كما يوضح الشكل (٦-٨)، تشارك صفّاً (وعموماً) مع ركائز أخرى. وبالتالي، فإن جميع الركائز في صف معين لها شيء مشترك - فهي تمتلك سمة مشتركة (عضوية في نفس الصف) لم تكن واضحة من موقع الرؤية السابق. إن مجرد تغيير زاوية الرؤية جعل شيئاً ما يتعلق بطبيعة الأشياء واضحاً كان في منظورنا الأولي غامضاً. دوران العوامل يشبه هذه



المحاكاة في أنه يحاول توفير “وجهة النظر” التي من خلالها يصبح الهيكل التنظيمي للبيانات - الطرق التي تشترك بها البنود في بعض الخصائص - واضحاً.



شكل (٦-٧): ترتيب منظم للركائز التي يتم رؤيتها من منظور يحجب ذلك الترتيب



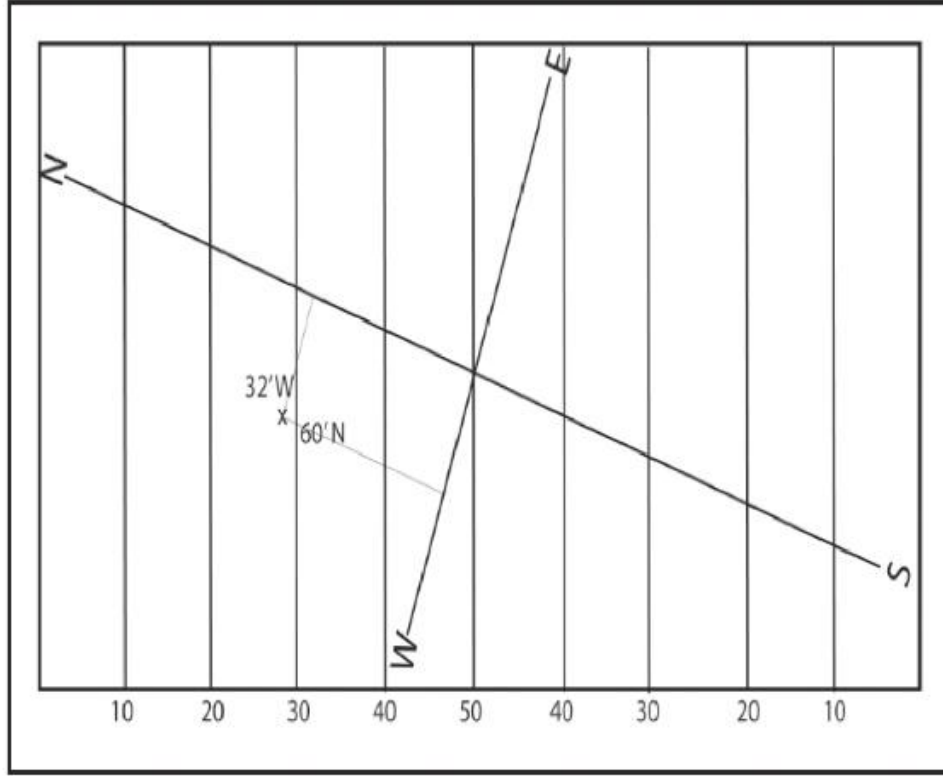
شكل (٦-٨): ترتيب منظم للركائز التي تتم رؤيتها من منظور يكشف عن ذلك الترتيب

#### التشبيه الثاني لتدوير العوامل:

تجدر الإشارة إلى أنه، مع العدد الصحيح من الخطوط العمودية المرجعية، يمكن للمرء تحديد موقع الأشياء بغض النظر عن كيفية توجيه هذه الخطوط المرجعية. مثال ثنائي الأبعاد مثل ملعب كرة قدم كبير فارغ يمكن استخدامه للتوضيح. لجعل هذا المشهد أكثر إثارة للاهتمام قليلاً، دعونا نفترض أن المدرب يستخدم مباراة فيديو، تم تصويرها من كاميرا علوية، لشرح اللعب لأعضاء الفريق. علاوة على ذلك، افترض أن المدرب قد قام بتجميد الصورة في لحظة مثيرة للاهتمام بشكل خاص أثناء اللعب لتوضيح نقطة معينة وأن هذه الصورة المجمدة يتم عرضها على شاشة كمبيوتر كبيرة تتيح للمدرب الرسم فوق تلك الصورة. يمكن للمدرب رسم خط مستقيم من

خلال صورة ملعب كرة القدم في أي اتجاه ثم وضع خط ثانوي عمودي على الأول. على سبيل المثال، قد يختار المدرب خطأً واحدًا يمتد تمامًا شمالًا وجنوبًا وآخر يمتد على وجه التحديد شرقًا وغربًا، والتي قد تسمح له بوصف أي موقع في الحقل باستخدام إحداثيات تحديد المواقع الشاملة استنادًا إلى خط العرض من الشمال إلى الجنوب وخط الطول من الشرق إلى الغرب. باستخدام هذين الخطين، يمكن للمدرب تحديد موقع اللاعب الذي يمتلك الكرة: «باستخدام إحداثيات تحديد المواقع الشاملة لوسط الملعب كمرجع، يكون حامل الكرة عند تقاطع نقطة محددة من خلال التقدم شمالًا من وسط الملعب لمسافة 60 قدمًا ثم المضي قدمًا غربًا 32 قدمًا إضافيًا». هذا سيحدد لاعب كرة القدم في مكان معين. يمكن للمدرب تحديد الموقع نفسه باستخدام تعليمات معدلة بشكل مناسب بناءً على أي خطين عموديين مرسومين في ملعب كرة القدم. لذلك، فإن اتجاه الخطوط هو أمر اعتباطي فيما يتعلق بكفاية وصف موقع معين. أي مجموعة من خطين عموديين لهما نفس الفائدة المعلوماتية في تحديد مكان معين مثل أي مجموعة أخرى.

بدلاً من ذلك، بدلاً من استخدام المحورين شمال - جنوب وشرق - غرب، يمكن للمدرب استخدام الخطوط الجانبية وخطوط المرمى للملعب كمحاور متعامدة أساسية. مثل أي مجموعة أخرى من الخطوط المتعامدة، يمكن استخدام هذه المحاور كمراجع لتحديد موقع أي نقطة في الملعب. ومع ذلك، فهي تتمتع بميزة الاستفادة من ميزة ذات مغزى في الملعب — تنظيمها كشبكة محددة بخطوط متعامدة (أي خطوط ساحة الملعب والخطوط الجانبية). وبالتالي، فإن إخبار شخص ما بأن حامل الكرة كان 4 أقدام داخل خط الـ 30 ياردة وحوالي 68 قدمًا من الخط الجانبي لملعب الفريق المضيف، قد تكون وسيلة أسهل لتحديد موقع هذا اللاعب من استخدام الخطوط المرجعية الممتدة بين الشمال والجنوب والشرق والغرب. النقطة المهمة هي أنه في حين أن أي مجموعة من المحاور المرجعية يمكنها تحديد موقع حامل الكرة بشكل لا لبس فيه، فإن المجموعة القائمة على الخصائص الطبيعية لملعب كرة القدم تبدو أكثر ملاءمةً وشموليةً (انظر الشكل رقم ٦-٩).



شكل (٦-٩): وصف لملاعب كرة قدم معلم بخطوط الميدان،  
والخطوط الجانبية، وإحداثيات بين الشمال - الجنوب والشرق - الغرب

الآن وقد رأينا كيف يمكن استخدام المتجهات أو الخطوط المرجعية لتحديد موقع مكاني، يجب أن نفحص كيفية ارتباط البنود بخطوط مرجعية مماثلة. ومع ذلك، قد يكون من الصعب التبديل بين التفكير في البنود ومحتواها من ناحية والتفكير في المتجهات المجردة التي تحدد المواقع من ناحية أخرى. في كثير من الأحيان، الأمثلة التي تربط البنود الأكثر مألوفية بتمثيلها المكاني على طول المتجهات يجب التغاضي عنها، لأن الأشياء التي نقيسها كعلماء اجتماع وسلوك لا تتناسب عادةً مع ترجمة مباشرة بالحدود المكانية. لسد هذه الفجوة، سأقدم مثالاً مختلفاً (غير حقيقي) استناداً إلى بنود يمكننا ربطها بالمواقع المادية.

التشبيه الثالث لتدوير العوامل:

امتداداً لمثال كرة القدم لدينا، سوف نفترض دراسة تستخدم بنود الاستبانة كوسيلة لوصف موقع لاعب كرة القدم (الشخص الذي يحمل الكرة) في الملعب. وبشكل أكثر تحديداً، دعنا نفترض أنه يُطلب من مجموعة من المُستطلّعين مشاهدة فيديو عن لعبة كرة القدم، وعند نقطة ما، يتم تجميد الصورة ويطلب من المجيبين الإجابة على سلسلة من ثمانية بنود باستخدام مقياس ليكرت المتدرج والذي يتراوح من 1 = لا أوافق بشدة إلى 6 = أوافق بشدة. هذه البنود كالتالي:

- A- حامل الكرة في موضع تهديف سهل.
- B- قد يتم إجبار حامل الكرة على الخروج من الحدود سريعاً.
- C- قد يُجبر حامل الكرة إلى منطقة النهاية الخاصة به.
- D- حامل الكرة لديه مساحة كبيرة لأي من الجانبين.
- E- إذا تم إيقاف حاملة الكرة الآن، شخص ما على الخطوط الجانبية قد يكون في خطر.
- F- أمام حامل الكرة طريق طويل قبل عبور خط المرمى.
- G- حامل الكرة داخل النطاق الميداني للمرمى.
- H- يجب أن يكون حامل الكرة حريصاً على عدم تجاوز الخط الجانبي.

قبل المتابعة مع هذا المثال، دعونا نتوقف للحظة بشكل أكثر رسمية للتعرف على كيفية دوران العوامل. عندما يكون لدينا عاملان، فإنه يمكن تحقيق الدوران بيانياً. ويمكن للمرء الحصول على الشكل الانتشاري scatterplot الذي يمثل قوة الارتباطات بين أزواج من البنود عن طريق القرب من هذه البنود وتشبعات loadings تلك البنود على متجهات مرجعية اعتباطية مقابلة لاثنتين من العوامل الغير مدارة. يمكن للمرء إجراء دوران رسومي من خلال تدوير تلك المحاور حتى تسقط على مجموعات البنود. بعد ذلك، يتم استخدام المحاور المرجعية لإنشاء إحداثيات تحدد الموضع الدقيق لكل بند ممثّل من حيث موقعه بالنسبة إلى الأطوال على امتداد المحورين. ويمكن بعد ذلك تحويل هذه الإحداثيات إلى تشبعات لكل بند على هذين العاملين. ولكن طرق الدوران غير الرسومية الآلية القائمة على المعايير الرياضية تستخدم على نطاق أوسع. على سبيل المثال، تسعى

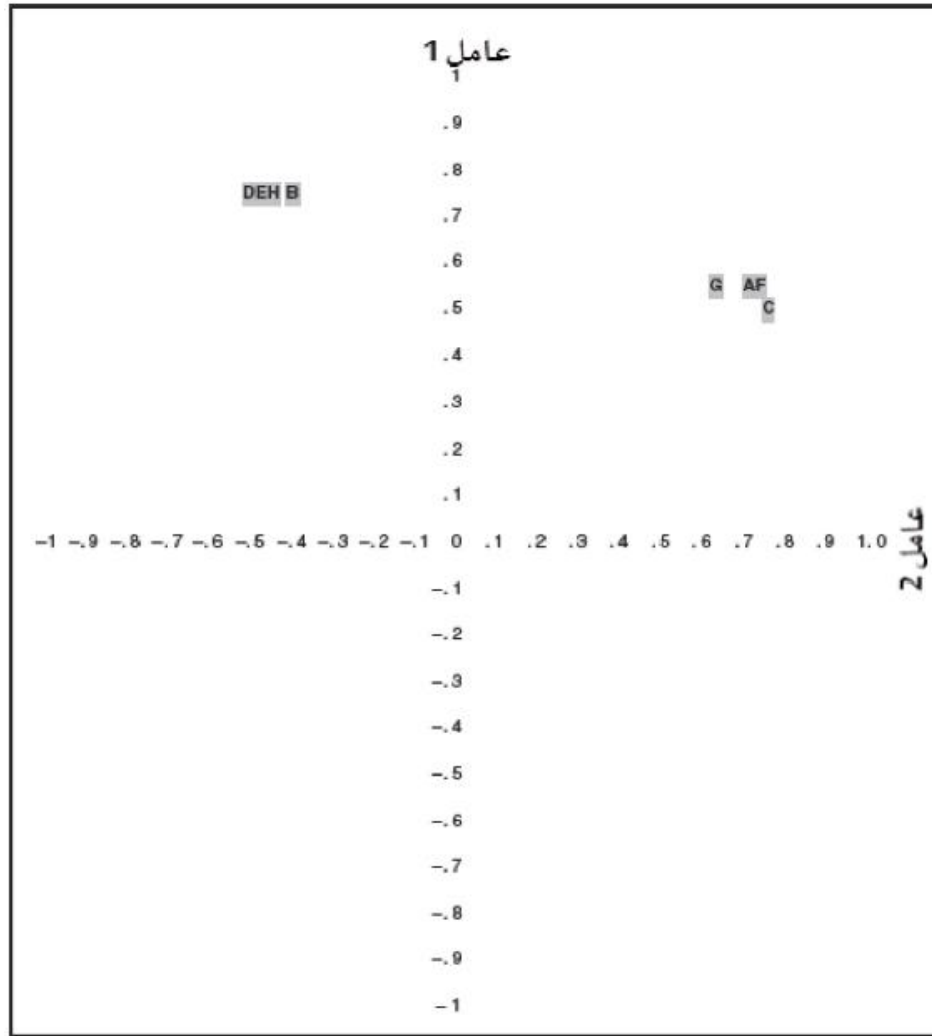
أساليب مثل التباين الأعظم Varimax إلى زيادة تباين التشبعات المربعة لكل بند. التشبعات هي عبارة عن ارتباطات بين كل بند وكل عامل بناءً على اتجاه المتجهات (أي الخطوط المرجعية) التي تحدد العوامل. وبناءً على الكيفية التي يتم بها توجيه المتجهات، ستتباين التشبعات، تمامًا كما تتباين القيم التي تحدد موقع لاعب كرة قدم في الملعب أو الإحداثيات التي تصف موضع نقطة ما على الشكل الانتشاري اعتمادًا على المحاور المرجعية المستخدمة. سيكون هذا التباين أكبر عندما تكون بعض التشبعات المربعة كبيرة نسبيًا والبعض الآخر صغيراً نسبياً. وبالتالي، فإن تلبية معيار تعظيم تباين التشبعات المربعة له تأثير في إنتاج مجموعة متفاوتة القيم القصوى من التشبعات لكل بند. العثور على اتجاه للمحاور المرجعية التي ستمنحنا تشبعات متفاوتة القيم القصوى والقيم الأكبر لتباينات التشبعات المربعة سيؤدي إلى إيجاد الدوران الذي من المحتمل أن يكون مفيداً. ما نسعى إليه، بالطبع، هو نوع معين من التفاوت بحيث يكون تشبعاً واحداً (أي، تشبع البند على العامل الأساسي) كبيراً في حين أن جميع التشبعات الأخرى لهذا البند (على سبيل المثال، التشبعات الخاصة به على العوامل الثانوية) صغيرة (من الناحية المثالية، قريبة من الصفر). عندما يتحقق ذلك، يمكننا وصف البند بأنه يتأثر بشكل أساسي بالعامل الفردي الذي يتشبع عليه بشكل كبير - وهي الحالة التي تدعى الهيكل البسيط. إن وصف بندٍ من خلال عامله الأساسي فقط يقوم بعمل جيد لالتقاط الطبيعة الأساسية للمعلومات التي يحملها البند. هذا هو هدف الدوران.

الآن، دعونا نعود إلى البنود التي تصف موقع حامل الكرة في الصورة المجمدة للعبة كرة القدم. لقد تم تحليل هذه البنود الثمانية (بناءً على بيانات مزيفة) لأغراض توضيحية. وكانت الجذور الكامنة الثلاثة الأولى 67.3 و 78.2 و 59.0. بناءً على الحقيقة أن القيمتين الأوليين أكبر بكثير من 0.1 بينما الثالثة أصغر بكثير من 0.1، وتم استخراج عاملين. حيث طلب من برنامج SAS إجراء دوران بأسلوب التباين الأعظم Varimax على العوامل. الشكلان رقمي (٦-١٠) و (٦-١١) يوضحان رسومات العوامل باستخدام برنامج SAS قبل وبعد الدوران، على التوالي. وهذه تحدد موقع البنود فيما يتعلق بالعاملين بناءً على تشبع كل بند على هذين العاملين.

يقوم SAS بطباعة رسومات بيانية خاصة للعوامل بخط يستخدم العرض المتساوي لكل حرف ولا يستطيع التقدير interpolate بين خطوط من النصوص. وبالتالي، لا يمكن تحديد موقع النقاط بالنسبة لبعضها البعض أو إلى متجهات العوامل بدقة تامة. ومع ذلك، حتى مع هذا القيد، فإن

الرسومات البيانية إيحائية تماماً. في الشكل رقم (٦-١٠)، يتم تجميع البنود (ممثلة بأحرف داكنة في مربعات مظلمة لرؤية أكثر وضوحاً) في مجموعتين، إحداهما تشتمل على البنود B و D و E و H والأخرى، البنود A و F و C و G. القرب بين البنود يشير إلى قوة الارتباط في هذا الشكل. لذلك، أي بندين يقعان بالقرب من بعضهما البعض مرتبطان بدرجة كبيرة أكثر من أي زوج بعيد. وبالتالي، يشير وجود مجموعتين في هذا الشكل إلى أن البنود الموجودة في كل مجموعة ترتبط ارتباطاً وثيقاً ببعضها البعض وأن البنود الموجودة في مجموعات مختلفة غير مرتبطة إلى حد كبير ببعضها البعض. المواقع بالنسبة إلى المحاور هي أيضاً مفيدة وذات صلة خاصة بمسألة الدوران. تقع مجموعة البنود A-F-C-G على بعد متساوٍ تقريباً من المتجهين الممثلين للعوامل. أي إن تشعباتها على كل عامل تقريباً بنفس الحجم. مجموعة البنود الأخرى، على الرغم من أنها أقرب إلى حد ما من المتجه الرأسي منها إلى المتجه الأفقي، فإنها لا تزال على مسافة من أي منهما. تحتوي البنود الموجودة في المجموعة الثانية على تشعبات أعلى على العامل 1، ولكن أيضاً تشعبات (سالبة) كبيرة إلى حد ما على العامل 2. وبالتالي، سيكون من الصعب القول إن أي من مجموعتي البنود يعكس بشكل أساسي مستوى واحداً فقط من العوامل.

في المقابل، في الشكل الذي تم إنشاؤه بعد أن تم تدوير العوامل، نلاحظ أن كلاً من المجموعتين قريبة جداً من متجه واحد فقط. المجموعة الأولى من البنود (B-D-E-H) لديها الآن تشعبات عالية (تقريباً من 8.0 إلى 9.0) على العامل 1 (يمثله المحور الرأسي) وتشعبات منخفضة على العامل 2 (أقل من 2.0). المجموعة الثانية (A-F-C-G) قريبة من العامل 2 (المحور الأفقي)، وكل البنود مجتمعة تمثل قيم عالية على العامل 2 (تقريباً في حدود 80.0 إلى 90.0) وقيم منخفضة على العامل 1 (حوالي 0.10).



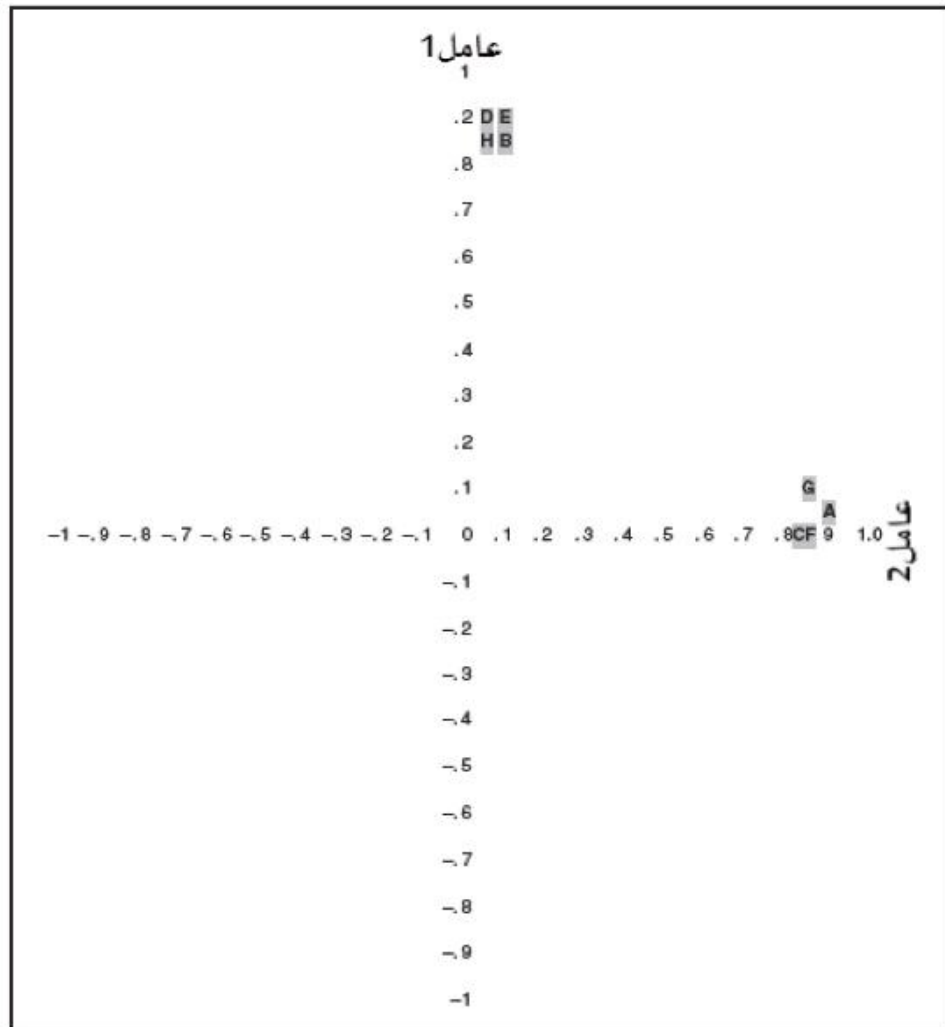
شكل (٦-١١): العناصر الممثلة بيانياً بواسطة التشبعات على عاملين قبل الدوران

لذلك، بعد الدوران، لدينا تقريب أفضل بكثير لهيكل بسيط من توزيع البنود على العوامل - أي، كل بند يتشعب بشكل أساسي على عامل واحد.

إذا نظرت بعناية إلى المواقع النسبية لمجموعتي البنود في الشكلين المختلفين، قد تكون قادرًا على التمييز على أنهما متشابهان إلى حد كبير (في حدود الرسم البياني لـ SAS). مجموعة البنود A-F-C-G تقع حوالي 90 درجة في اتجاه عقارب الساعة من المجموعة B-D-E-H في كلا الشكلين. هذا كما يجب أن يكون، لأن الدوران لا يغير العلاقات بين المتغيرات. لو كانت البنود



مرسومة على ورقة شفافة والعوامل على ورقة شفافة أخرى، فإنه يمكنك تدوير الورقة المحتوية على البنود من الموضع الموضح في الشكل رقم (٦-١٠) إلى الموضع الموضح في الشكل رقم (٦-١١) الخاصة بالعوامل.



شكل (٦-١١): العناصر الممثلة بيانياً بواسطة التشعبات على عاملين بعد الدوران

(حقيقة أن المحاور الأفقية والعمودية لها مقاييس مختلفة قليلاً بسبب عدم الدقة في كيفية عرض SAS للرسوم البيانية، قد تجعل مسافات المجموعات من الأصل أقل تشابهاً في الشكلين مما هي عليه بالفعل عددياً). في الحقيقة، إن الدوران يعيد فعلاً توجيه العوامل، ولكن بدلاً من عرض محاور العامل بزوايا غريبة، تقوم رسومات ما بعد الدوران بشكل أساسي بتدوير مواقع البنود في اتجاه عقارب الساعة بدلاً من تدوير المحاور في اتجاه عكس عقارب الساعة.

بعد تدوير العوامل، أصبح من الممكن الآن تفسير كل منها من خلال النظر في ماذا تشترك البنود التي حققت تشعبات عالية على كل من العوامل. البنود التي تشكل العامل الأول جميعها تتعلق بالقرب من الخطوط الجانبية للملعب، وتلك التي تشكل العامل الثاني تتعلق بالقرب من خطوط المرمى. وبالتالي، قد نفسر العامل 1 على أنه القرب من الخطوط الجانبية والعامل 2 على أنه القرب من خطوط المرمى. لقد سمح لنا الدوران بتجميع البنود بطريقة سهلت التعرف على السمة العامة المشتركة بين البنود داخل كل مجموعة.

باختصار، إن أهم النقاط التي تدل عليها هذه الأمثلة حول دوران العوامل هي كما يلي: (أ) معرفة العدد الصحيح للعوامل (عندما يكون هناك أكثر من عامل واحد)، رغم أنه ضروري، لا يكشف عن طبيعة تلك العوامل. عملية استخراج العوامل الأولية تتناول فقط عدد العوامل المطلوبة، وليس طبيعة تلك العوامل بأي معنى جوهري. (ب) في حالة غياب دوران العوامل المتعددة، فإن أي نمط ذي معنى وقابل للتفسير بين البنود يمكن حجبها. الحل الأولي لاستخلاص العوامل يتجاهل محتوى البنود، وعادةً ما تكشف تجمعات البنود بناءً على عوامل غير مدارة القليل عن تلك البنود. ويوضح تصوير الركائز من منظور غير ملائم ومن ثم منظور ملائم هذه النقطة. (ج) يمكن لأي حل تحليلي يشتمل على العدد الصحيح من العوامل تحديد العلاقات بين البنود (أو، إذا وصفناها مكانياً، أي مواقع البنود بالنسبة لبعضها البعض) بالإضافة إلى غيرها. تعمل المحاور الشمالية - الجنوبية والشرقية - الغربية تمامًا كما تعمل محاور المرمى - المرمى والخطوط الجانبية - الخطوط الجانبية لتحديد أي موقع في الملعب. (د) الغرض من تدوير العوامل هو إيجاد اتجاه معين للمحاور المرجعية التي تساعدنا على فهم البنود بعبارات بسيطة. يتم تحقيق ذلك عندما يمكن شرح معظم محتويات البنود باستخدام أقل عدد ممكن من المتغيرات - من الناحية المثالية، واحد فقط. (هـ) يمكن أن يسهل نمط العوامل المدارة تحديد ما هو مشترك بين البنود داخل عامل ما وبالتالي

استنتاج ما هو العامل السببي الأساسي الذي يحدد كيفية الإجابة على البنود. كانت المجموعتان من البنود المتعلقة بموقع حامل الكرة في لعبة كرة القدم مختلفة بشكل واضح. مجموعة واحدة تتعلق فقط بالقرب من الخطوط الجانبية، في حين أن المجموعة الأخرى تتعلق فقط بالقرب من خطوط المرمى. يمكننا أن نستنتج، على الأقل مؤقتاً، أن تصورات القرب من الحدود المختلفة لملاعب كرة القدم تحدد كيفية تقييم المجيبين لمجموعتي البنود. بالطبع، ستكون هناك حاجة إلى معلومات تحقق إضافية لدعم هذا الاستدلال.

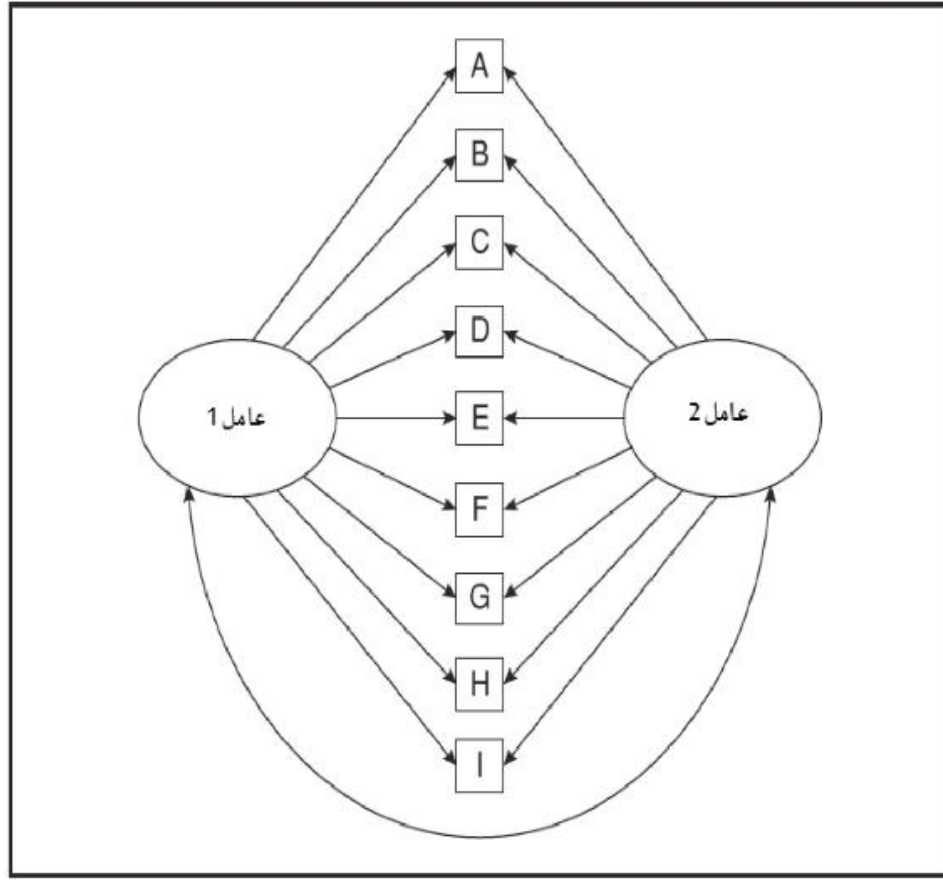
#### التدوير المتعامد مقابل التدوير المائل:

تستند جميع الأمثلة حتى الآن إلى خطوط مرجعية متعامدة مع بعضها البعض. هذا يتوافق مع العوامل التي هي مستقلة إحصائياً عن بعضها البعض (أي غير مترابطة). مثل هذه العوامل توصف بأنها متعامدة orthogonal. لا يوفر الموقع على طول خط (محور) ما أي معلومات تتعلق بالمعلومات الموجودة على طول الخط الآخر عندما يكون الخطان متعامدين. على سبيل المثال، معرفة مدى بعد أي شخص شمالاً لا تعطي مؤشراً على معرفة مدى بعده غرباً، لأن هذين الاتجاهين متعامدان مع بعضهما البعض. وبالمثل، معرفة مدى بُعد اللاعب عن خط المرمى لا تعطي مؤشراً على قربه من الخط الجانبي. ومع ذلك، من الممكن السماح للعوامل بالارتباط، وبالتالي، فإن المحاور التي تمثلها بشكل بياني تكون غير متعامدة. لنفترض، على سبيل المثال، أننا اخترنا تحديد مواقع في ملعب كرة القدم كما هو موضح في الشكل (6-9) باستخدام خط يمتد من منطقة النهاية إلى منطقة النهاية كمحور وخط يمتد شرقاً وغرباً مثل الآخر. التحرك في اتجاه غربي يعني أيضاً التحرك نحو إحدى مناطق النهاية. أي أنه لا يمكن للمرء أن يتحرك مباشرة شرقاً أو غرباً دون تغيير مسافته عن خطوط المرمى. البعدان يرتبطان إلى حد ما.

بالمثل، يمكن أن يسمح دوران العوامل للمحاور المرجعية (والعوامل التي تتوافق معها) لتكون مرتبطة، وبالتالي، لا تكون متعامدة مكانياً. يشار إلى هذا الدوران على أنه مائل oblique بدلاً من المتعامد. قد يكون الدوران المائل مفيداً عندما يُعتقد أن المتغيرات الكامنة ترتبط ببعضها البعض. يتطلب هدف الهيكل البسيط البنود التي يمكن تصنيفها بشكل مفيد فيما يتعلق بفئة واحدة فقط. أي، يجب أن يكون كل بند "عن شيء واحد فقط، وبالتالي، يتشعب load على عامل واحد فقط. إذا كانت المتغيرات مرتبطة إلى حد ما ولكن العوامل التي تمثلها أُجبرت إلى أن تكون مستقلة

تمامًا بسبب القيود التي تفرضها طريقة تحليل العوامل، فقد يكون من المستحيل تحقيق هذا الهدف. أي إن أكثر من عامل قد يرتبط ببعض أو كل البنود بسبب الارتباط بين العوامل؛ نحن محدودون في قدرتنا على التقريب للهيكल البسيط (انظر الشكل رقم ٦-١٢).

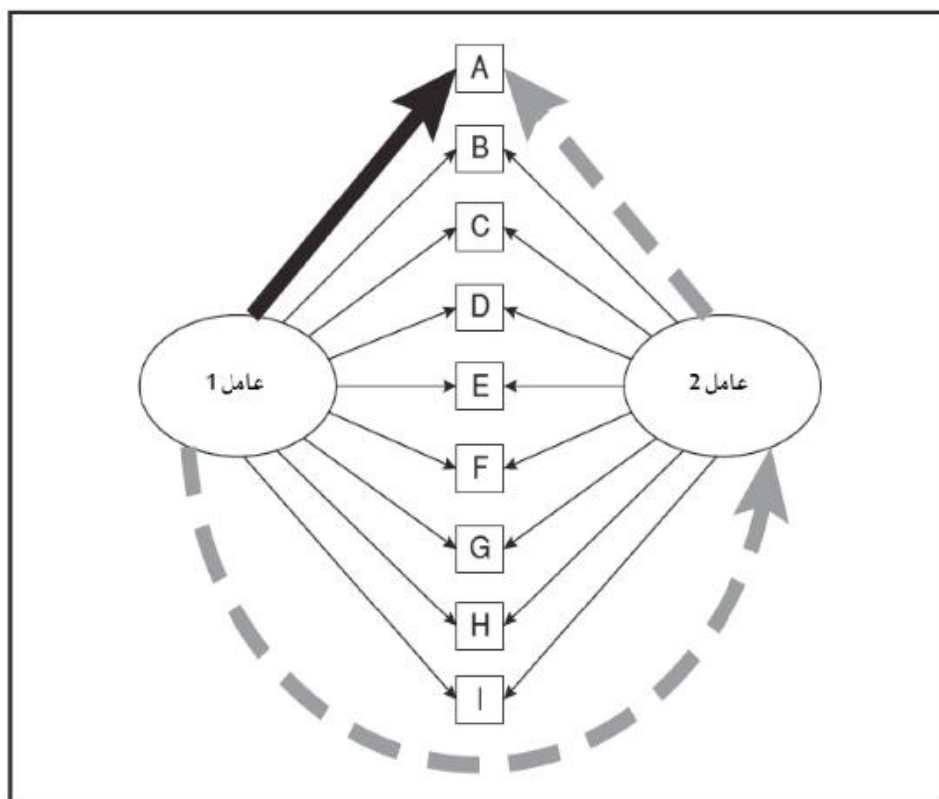
بالعودة إلى مثال خصائص زميل العمل السابق، إذا كان “الاجتهاد” و “الاعتمادية” مترابطين حقًا، فمن المحتمل أن يتقاسم بند عن أحدهما بعض التباين مع الآخر أيضًا. الأشخاص الأكثر اجتهادًا سيكونون أيضًا أكثر اعتمادًا إلى حد ما (بدرجة أكبر أو أقل استنادًا إلى قوة الارتباط بين السمتين). إن فرض حل متعامد على البيانات التي تعكس هذين المتغيرين سيجعل من الصعب للغاية على البند أن يتشعب بشكل قوي على متغير واحد وبشكل ضعيف على الآخر. ومع ذلك، إذا سمح للعاملين بأن يكونا مترابطين إلى حد ما، تكون الحالة مشابهة تقريبًا للاستنتاج التالي: “الاجتهاد” و “الاعتمادية” يُفهم أنهما يرتبطان ببعضهما البعض. وقد تم التعامل مع هذه الحقيقة من خلال السماح للعوامل بالارتباط. الآن، إلى جانب ذلك، إلى أي من هذه العوامل يرتبط البند المعني بشدة؟ إذا لم يتم إجبار المحاور التي تمثل عامل “الاجتهاد” وعامل “الاعتمادية” على أن تكون متعامدة، فمع زيادة الدرجات على أحد المتغيرات، يمكن استيعاب درجة ما من الزيادة المصاحبة في المتغير الآخر.



شكل (٦-١٢): نموذج بعاملين مع السماح للعاملين بالارتباط

إذا كان الواقع يفرض أن البند الذي يتأثر بشكل كبير بـ «الاجتهاد» يتأثر أيضاً بـ «الاعتمادية»، فإن السماح للعوامل التي تمثل تلك المفاهيم البنائية بالترابط يمكن أن يستوعب تلك الحقيقة بشكل أكثر دقة. قد يكون البند الذي يرتبط ارتباطاً وثيقاً بـ «الاجتهاد» ولكنه يرتبط بشكل معتدل بـ «الاعتمادية» قادراً على التشبع بقوة على أول تلك العوامل وبشكل ضعيف على الثاني إذا كانت العوامل غير مقيدة بالاستقلال الإحصائي. بمعنى، حقيقة أن هذا البند له تأثير سببي ناشئ عن «الاعتمادية»، يتم استيعابه من خلال السماح للعاملين بالترابط مع بعضهما البعض. يوفر هذا مساراً غير مباشر من «الاعتمادية» إلى البند من خلال «الاجتهاد» (انظر الشكل رقم ٦-١٣)، مما يزيل الحاجة إلى ارتباط البند مباشرة بـ «الاعتمادية» وبالتالي التشبع المتقاطع على كلا العاملين.

ما يتم فقدانه عندما يتم تدوير العوامل بشكل مائل (غير متعامد) هو أناقة وبساطة الأبعاد غير المترابطة. إحدى المزايا الجذابة للعوامل غير المترابطة هي أن تأثير العوامل مجتمعة عبارة عن حاصل المجموع البسيط لتأثيراتها منفصلة. مقدار المعلومات في قيمة بند معين الذي يفسره أحد العوامل يمكن إضافته إلى مقدار المعلومات الذي يفسره عامل آخر للحصول على إجمالي مقدار المعلومات المفسرة بالعاملين مع بعضهما البعض. في العوامل المائلة، الحال ليس كذلك.



شكل (٦-١٣): نظرًا للارتباطات بين العوامل، يؤثر العامل 1 على البند A بشكل مباشر (المسار الصلب الداكن) وغير مباشر (المسار المتقطع الفاتح)

ولأن العاملين مترابطان، يوجد تكرار في المعلومات الواردة في هذين العاملين. بالنسبة إلى بند يتعلق بكل من العاملين «الاجتهاد» و«الاعتمادية»، فإن مقدار التباين الذي يفسره هذان العاملان معًا أقل من مجموع تباين العوامل منفصلة. بعض المعلومات، وربما قدرٌ كبيرٌ منها، التي يستأثر بها أحد العوامل تتداخل مع المعلومات التي يستأثر بها العامل الآخر. محصلة المجموع البسيط

للمعلومات ستتضمن المعلومات المتداخلة مرتين، والتي لن تعكس بدقة التأثير الكلي للعوامل على ذلك البند.

ومن المضاعفات الأخرى للعوامل المترابطة التعقيد المضاف للعلاقات السببية بين البنود والعوامل. عندما تكون العوامل مستقلة، تكون العلاقة الوحيدة بين عامل وبند معين مباشرة. ستؤدي التغييرات في مستوى العامل إلى تغييرات في البند على امتداد مسار سببي مباشر واحد. غير أن هذا ليس هو الحال عندما تكون العوامل مترابطة. إذا كان (كما هو موضح سابقاً عند مناقشة بند متأثر بـ “الاجتهاد” و “الاعتمادية”) يؤثر كلاً من عاملين افتراضيين على البند A، على سبيل المثال، وكان العاملان مترابطين، فإن العاملين يمارسان تأثيراً غير مباشر وكذلك مباشر على البند A. بمعنى، يمكن للعامل 1 التأثير على العامل 2، ومن خلال العامل 2، التأثير بشكل غير مباشر على البند A. هذا بالإضافة إلى التأثير المباشر للعامل 1 على ذلك البند.

بالطبع، وبعملية موازية، يمكن للعامل 2 أيضاً التأثير على البند ليس فقط بشكل مباشر ولكن أيضاً بشكل غير مباشر من خلال علاقته بالعامل 1. نفس النوع من التأثير المباشر وغير المباشر للعوامل ينطبق على جميع البنود الأخرى. نتيجة لذلك، عند الحديث عن العلاقة بين بند وعامل فإنه عادةً ينبغي أن تكون هذه العلاقة مؤهلة بشكل صريح إما لإدراج أو استبعاد مثل هذه التأثيرات غير المباشرة. خلاف ذلك، يوجد غموض، وبالتالي ارتباك محتمل.

#### اختيار نوع التدوير:

كمسألة عملية، يجب أن يعتمد الاختيار بين الدوران المتعامد والدوران المائل على واحد أو أكثر من الاعتبارات. من بين هذه الاعتبارات كيف ينظر المرء إلى المفاهيم التي تمثلها العوامل. إذا كانت النظرية تشير بقوة إلى مفاهيم مترابطة، فربما يكون من المنطقي أن يتبع النهج التحليلي للعوامل (على وجه التحديد، الدوران) ذلك. وبالتالي، إذا كان تحليل العناصر المتعلقة بـ “الاجتهاد” و “الاعتمادية”، فإن السماح للعوامل بالارتباط من شأنه أن يناسب إدراكنا بما تنطوي عليه هذه المفاهيم. عوضاً عن ذلك، قد تقترح النظرية عوامل متعامدة. على سبيل المثال، قد تكون “الاعتمادية” و “المرح” أكثر استقلالية، وبالتالي قد تتطلب حلاً متعامداً. عندما لا تقدم النظرية دليلاً قوياً، مثلما يمثل مقياس قيد التطوير مفاهيم لم تتم دراستها من قبل، فإن حجم الارتباطات بين العوامل قد يكون بمثابة دليل. على وجه التحديد، يمكن تحديد الدوران المائل ودراسة قيم

الارتباطات الناتجة بين العوامل. إذا كانت قيم هذه الارتباطات صغيرة جداً (على سبيل المثال، أقل من 15.0)، فقد يختار محلل البيانات الدوران المتعامد. هذا يقدم حلاً وسطاً لتقريب الهيكل البسيط ولكن ينتج عنه نموذج أبسط. على سبيل المثال، قد تُظهر بعض البنود تشبعات ثانوية (بمعنى، تشبعات على عامل آخر غير تلك التي تنتسج عليها بقوة أكبر) والتي تزداد بشكل طفيف بالنسبة للحل المائل ولكنها لا تزال صغيرة بشكل كافٍ لربط كل بند بشكل غامض مع عامل واحد فقط. وبالتالي، قد تكون تشبعات بند معين على ثلاثة عوامل تم تدويرها بشكل مائل هي 78.0 و 16.0 و 05.0. عند اختيار الحل المتعامد، قد تكون التشبعات هي 77.0 و 19.0 و 11.0. على الرغم من أن النموذج الثاني يبتعد أكثر بقليل من الأول عن الهيكل البسيط، إلا أنه لا يزال من الممكن ربط البند المعني بشكل لا لبس فيه بالعامل الأول. وبالتالي، تمت التضحية بالقليل في سبيل اختيار النموذج الأبسط (أي المتعامد) في هذه الحالة. إذا كانت العوامل مرتبطة بدرجة أكبر، فإن اختيار الحل المائل قد يؤدي إلى تحسن كبير في التقريب للهيكل البسيط. على سبيل المثال، قد يتضاءل تشبع ثانوي مقداره 40.0 تم الحصول عليه بتدوير متعامد إلى 15.0 بحل مائل. لن يكون هذا هو الحال دائماً، ومع ذلك، فإن دراسة الفرق بين طريقتي الدوران فقط يمكن أن تشير بشكل قاطع إلى أي مدى تختلف في تقريب الهيكل البسيط.

المسألة العملية الأخيرة تتعلق بحجم الارتباط بين عاملين وكم يجب أن يكون حجم الارتباط قبل دمج العاملين في عامل واحد أكبر. لا توجد إجابة بسيطة على هذا السؤال لأن العلاقات بين البنود والعوامل تحتاج أيضاً إلى النظر فيها. ومع ذلك، في بعض الحالات، قد يكشف الدوران المائل أنه حتى عندما يكون هناك عاملان يرتبطان ارتباطاً كبيراً، فإن بعض البنود لها تشبعات كبيرة على كليهما. في هذه الحالة، قد يكون من المنطقي استخلاص عوامل أقل لمعرفة ما إذا كان العاملان المرتبطان إلى حد كبير يمكن دمجهما في عامل واحد. قد يكون، على سبيل المثال، جيداً أن تدعم بيانات الحياة الواقعية عاملاً واحداً يجمع بين البنود المتعلقة بـ “الاجتهاد” و “الاعتمادية” بدلاً من الفصل بينهما.

#### النماذج العاملة الثنائية والهرمية:

تقريباً، تفترض جميع أساليب تطوير المقاييس التي تمت مناقشتها في هذا الكتاب أن البنود المكونة للمقياس أحادية البعد. التحليل العاملي هو الأداة المفضلة لتحديد البنود المرشحة المحتملة



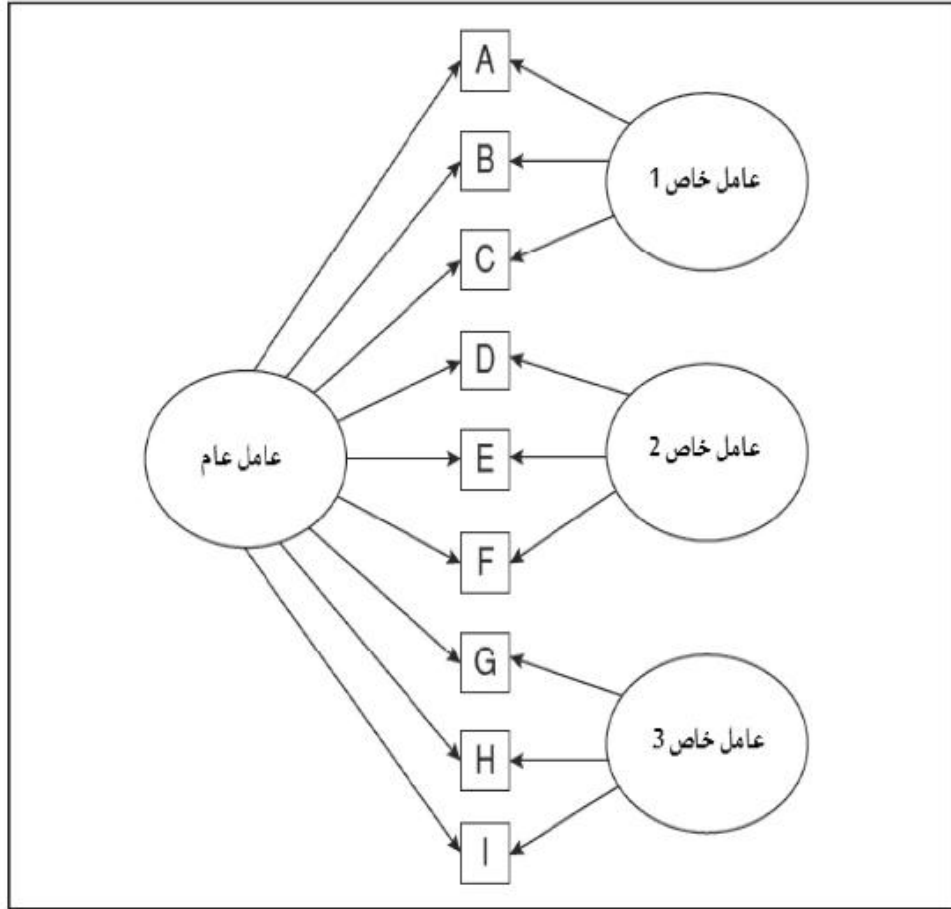
التي ستناسب هذا الافتراض. على الرغم من أن المجموعة الكاملة من البنود التي يهتم بها الباحث قد تكون متعددة الأبعاد، إذا كان بإمكانه تحديد مجموعات فرعية أحادية البعد، فإنه عندئذٍ يمكن معاملة كل من هذه المجموعات الفرعية كمقياس منفصل وأن العمل على تطوير أداة قياس (أو ربما بمزيد من الدقة، مجموعة من أدوات القياس ذات الصلة) يمكن أن يستمر.

قبل هذه النقطة، كانت مناقشتنا للتحليل العاملي حول النموذج التحليلي للعوامل المتعددة الأبعاد الأكثر شيوعاً، أي نموذج قائم على هيكل بسيط. تذكر أنه تم وصف الهيكل البسيط بأنه موجود عندما يكون لكل بند تشعب قوي على عامل واحد فقط (والذي أشرنا إليه كعامل أساسي) وتشعبات ضعيفة على جميع العوامل الأخرى (أي الثانوية). وبإعادة الصياغة من منظور العوامل بدلاً من البنود، بالنسبة لمفهوم الهيكل البسيط، يعد كل عامل محددًا قويًا لبعض البنود (أي البنود التي نعتبرها مكونًا لذلك العامل) ومحددًا ضعيفًا للبنود المتبقية. نحن على وشك النظر في بدائل لنماذج الهيكل البسيط، ولكن نحتاج بعض التوضيح أولاً. النموذجان الموصوفان في الأقسام التالية هما، مرة أخرى، نماذج متعددة الأبعاد (أي متعددة العوامل)، لكن العلاقات بين البنود والعوامل لا تتوافق مع الهيكل البسيط مثل النماذج السابقة. مقارنةً بنهج الهيكل البسيط، تمثل هذه النماذج طريقة مختلفة للتفكير في العوامل المتعلقة بالبنود. كما سنرى، بالنسبة لهذه النماذج، تتأثر البنود بما يمكن أن نسميه بالعامل الشامل global factor والعامل الموضعي local factor. وسنقوم في النهاية بتحسين هذه المصطلحات بحيث تتناسب مع كل نموذج من النماذج التي نناقشها، ولكن هذه المصطلحات العامة ستخدم غرضنا في الوقت الحالي وستسمح لنا بالتمييز بين هذين النموذجين كثنائي عن تلك التي تطرقنا لها سابقاً.

لماذا نحتاج إلى نماذج تحليلية للعوامل غير تلك القائمة على الهيكل البسيط؟ بعض مجموعات البنود ببساطة لا تتوافق مع الهيكل البسيط. قد تُظهر عاملاً أولياً (شاملاً) كبيراً ولكنها تظهر أيضاً أدلة على مجموعات من البنود تشير إلى مكونات فرعية متماسكة ومتجانسة (عوامل موضعية). بمعنى أنه يبدو أن البنود لها عضوية مزدوجة، في العامل الأول الذي تشترك فيه جميعها وأيضاً في عوامل إضافية تشترك فيها مجموعات فرعية من البنود. السؤال الذي يطرح نفسه هو ما إذا كانت هذه البنود بأي حال من الأحوال أحادية البعد أم أنها متعددة الأبعاد لا محالة. بالمعنى الدقيق للكلمة، لا يمكن وصف البنود على أنها أحادية البعد استناداً إلى الدليل بأن النموذج العاملي بسيط الهيكل سينتج. ومع ذلك، إذا تمكناً بطريقة ما من وضع هذا العامل الأول الشامل

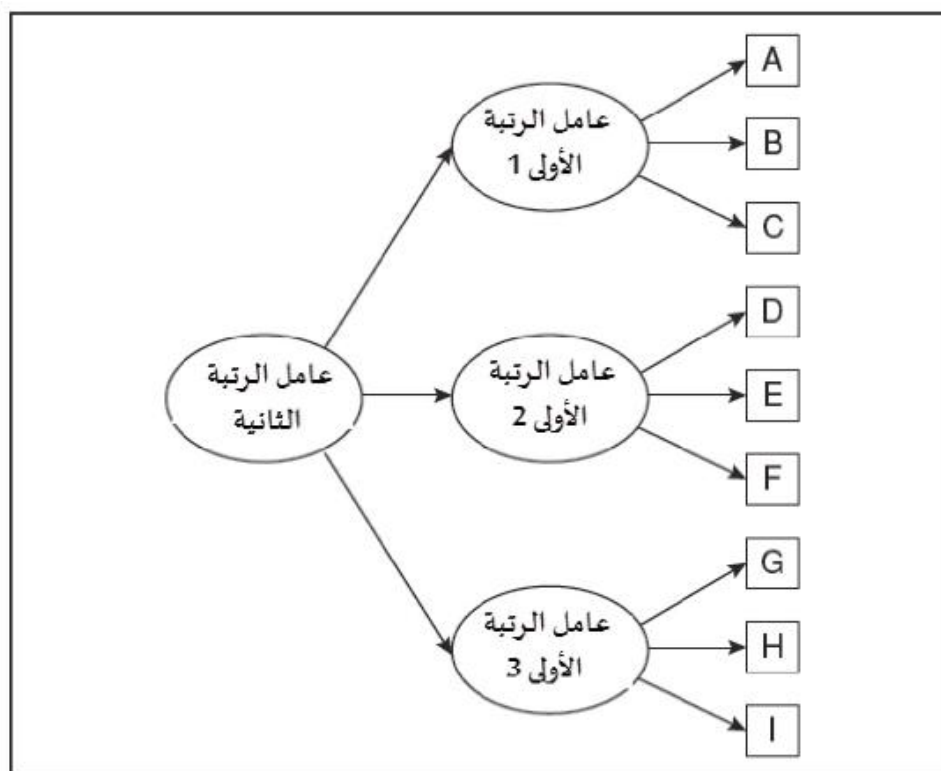
جانباً، فإن علاقات البنود بالعوامل الموضعية قد تبدو إلى حد كبير مثل المتغيرات الكامنة المعروفة بمجموعات فرعية من البنود التي يبدو أن لها هيكلًا بسيطاً وبالتالي فهي أحادية البعد داخل المجموعات الفرعية. ولكن كيف يمكننا تجاوز مشكلة العضوية المزدوجة والتغلب على نواة الهيكل البسيط؟ النماذج التي تم وصفها هنا قد توفر الوسائل لذلك. وقد قدم (Reise, Moore, Haviland, 2010) ملخصاً ممتازاً لهذه النماذج، وتمت الاستفادة من رؤاهم بشكل كبير فيما سيأتي.

سننظر في نوعين من النماذج غير ذات الهيكل البسيط. نموذج ثنائي العوامل model bifactor ويحدد عاملاً عاماً (اسم للعامل الشامل في سياق النماذج ثنائية العوامل) يساهم في التباين في جميع البنود ويحدد أيضاً مجموعة من العوامل الخاصة factors specific غير المترابطة (مصطلح العوامل الموضعية داخل النموذج ثنائي العوامل) التي تساهم في التباين في مجموعات فرعية من البنود (انظر الشكل رقم ٦-١٤). وبالتالي، في المثال الموضح، يتلقى كل بند تبايناً من العامل العام بالإضافة إلى عامل واحد خاص فقط. باستخدام النموذج ثنائي العوامل، لدينا تشكيلة من المتغيرات التي تشبه أحادية البعد على مستوى السمات الخاصة على الرغم من وجود سمة عامة. إذا تم تطبيق النهج التحليلي العاملي التقليدي بسيط الهيكل على البيانات التي تتناسب مع نموذج ثنائي العوامل، فإن البنود تميل إلى التشبع المتقاطع cross-load على العامل الأول (ملتقطة في المقام الأول نفس المتغير الكامن الذي يقوم به العامل العام في نموذج ثنائي العوامل) وسمة أو أكثر من السمات الخاصة. في هذا النهج، يؤدي حساب العامل العام بشكل أساسي إلى تحديد التباين المشترك بين جميع البنود. وما تبقى هي أكثر مصادر التباين محدودية (العوامل الخاصة specific factors). قد تظهر البنود الموجودة داخل كل من هذه العوامل أحادية البعد الأساسية. بمعنى أن كل بند قد يتم تشبعه بقوة على عامل واحد خاص فقط، وقد يحتوي كل عامل من العوامل الخاصة على بنود يمكن تفسيرها بسهولة على أنها تمثل مفاهيم بنائية constructs متميزة ومتماسكة.



شكل (٦-١٤): نموذج ثنائي العوامل

من الواضح أن النموذج ثنائي العوامل مناسب لمجموعات البنود التي لها متغير كامن شامل (أي، عام) وموضعي (أي، خاص) تمارس تأثيراً عليها. وينطبق هذا أيضاً على النهج التحليلي العاملية الآخر الذي قد يبدو مشابهاً للنموذج ثنائي العوامل للوهلة الأولى، وهو النموذج الهرمي hierarchical model. في النموذج الهرمي (انظر الشكل رقم ٦-١٥)، تتدرج مجموعة من العوامل الموضعية من الرتبة الأولى ضمن عامل شامل، أو أعلى رتبة، والتي تحدد المتغيرات الكامنة التي تمثلها عوامل الرتبة الأولى. على الرغم من أن المناقشة التالية تفترض وجود نموذج ذي مستويين فقط (أي عوامل متعددة من الرتبة الأولى وعامل واحد من الرتبة الأعلى)، إلا أنه من الناحية النظرية يمكن أن يمتد التسلسل الهرمي ليشمل عوامل من الرتبة الثانية كأسباب لعوامل الرتبة الأولى، وعوامل من الرتبة الثالثة كأسباب لعوامل من الرتبة الثانية، وهلم جرا.



شكل (٦-١٥): النموذج الهرمي

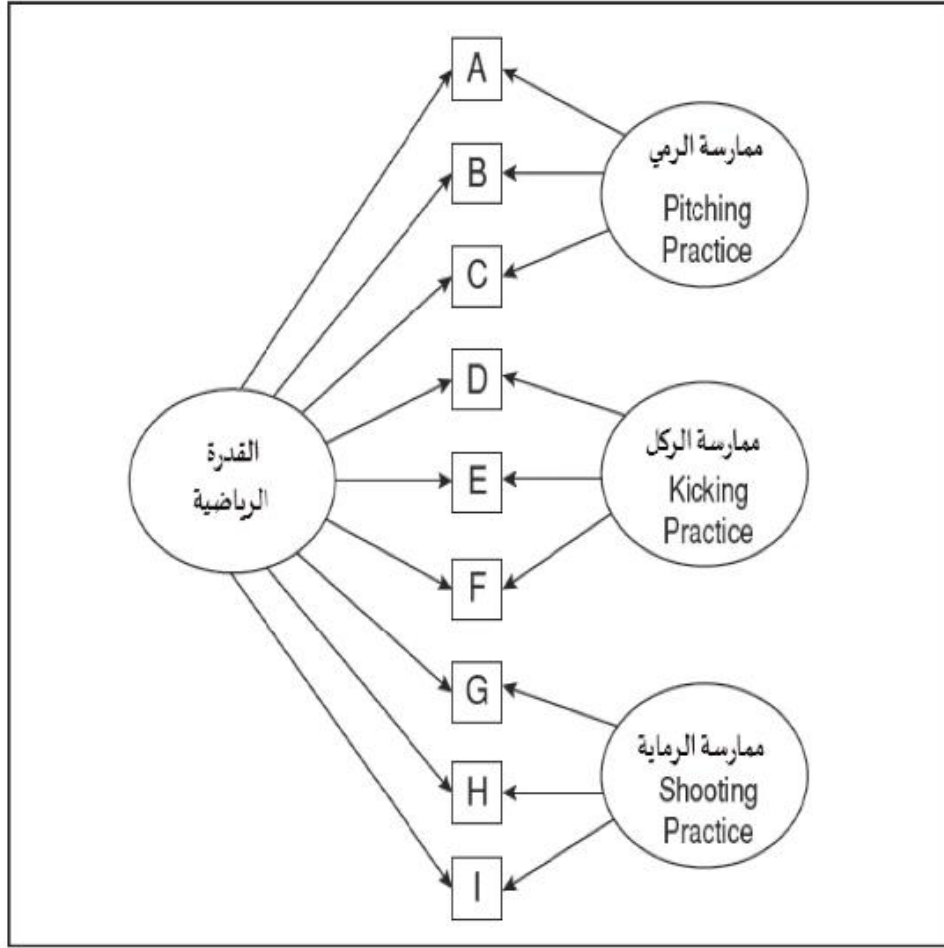
على الرغم من أن النماذج الهرمية والنماذج ثنائية العوامل قد تبدو متشابهة، حيث إن كليهما ينطوي على عوامل شاملة وموضعية، إلا أن هناك فرقاً مفاهيمياً مهماً؛ إنهما يختلفان في مفهوم السببية. في النموذج ثنائي العوامل، يتم تحديد قيمة كل بند بشكل مباشر من خلال كل من العامل الخاص والعامل العام، بينما في النموذج الهرمي، يكون عامل الرتبة الأعلى سبباً لعوامل الرتبة الأولى، والتي هي بدورها تحدد قيم البنود. ليس لعامل الرتبة الأعلى أي تأثير سببي مباشر على قيم البنود، ولكن كل هذا التأثير يتم بواسطة عوامل الرتبة الأولى (Reise et al., 2010). وبالتالي، بينما تكون البنود هي مؤشرات للعامل العام في النموذج ثنائي العوامل، فإنه في النموذج الهرمي، تكون عوامل الرتبة الأولى هي المؤشرات (الكامنة) لعامل الرتبة الأعلى. بالنسبة للنموذج الهرمي من النوع الموضح في الشكل رقم (٦-١٥)، فإن التحليل الذي يؤدي إلى عامل الرتبة الأعلى سيتضمن خطوتين مفاهيميتين: أولاً الحصول على الدرجات المناسبة من خلال التحليل العاملي لعوامل الرتبة الأولى (على سبيل المثال، لكل عامل من الرتبة الأولى، متوسط قيمة الدرجات التي يتم تشبعها عليه) ثم إجراء التحليل العاملي على درجات عوامل الرتبة الأولى تلك للحصول على تشبعاتها على عامل الرتبة الأعلى.

هناك مثالان مبسطان، واحد لكل من هذه النماذج التحليلية العاملية الجديدة، سيتم تناولهما مما قد يجعل هذه النماذج أكثر وضوحاً. حيث يمكننا اعتبار ما قد يكون نموذجاً تحليلياً عاملياً مناسباً؛ أولاً لمجموعة من القدرات الرياضية ومن ثم بالنسبة للآراء الاجتماعية والسياسية.

قد يكون لدى الشخص مهارات رياضية عامة مثل القوة وخفة الحركة. بالإضافة إلى ذلك، قد يكون لدى هذا الشخص مهارات خاصة قائمة على الممارسة ذات صلة بالأنشطة الرياضية الفردية. قد تشمل هذه القدرة على رمي كرة سريعة عبر لوحة المنزل plate home في لعبة البيسبول، أو ركل كرة القدم من خلال الأعمدة القائمة uprights في كرة القدم الأمريكية، أو الإسقاط المباشر لكرة السلة في الهدف من منتصف الملعب. هذه المهارات الأخيرة يتم تحديدها جزئياً من خلال القدرة الرياضية العامة ولكن يتم تحديدها أيضاً إلى حد كبير من خلال ممارسة المهارة الخاصة. وبالتالي، لا يمكن بالضرورة لأي شخص لديه قوة جيدة وخفة في الحركة أن يرمي كرات سريعة بسرعة 97 ميلاً في الساعة التي كانت باستمرار داخل الزاوية تماماً في منطقة ضرب كرة البيسبول. لذلك، إذا كان لدينا قدرٌ ما من مجموعة واسعة من المهارات الخاصة مثل

قدرة الشخص على رمي ضربات الكرة السريعة، فقد يكون من الأفضل وضع نماذج للمؤشرات المقابلة لتلك المجموعة من المهارات باستخدام نموذج ثنائي العوامل. هذا النموذج سيمثل كل مهارة كمهارة متأثرة بعامل القدرة الرياضية العام وعامل الممارسة الخاص بالمهارات. وبالتالي، فإن قيم مؤشر رمي ضربات الكرة السريعة ستتأثر بشكل مشترك بـ "القدرة الرياضية العامة" و "ممارسة الرمي". لاحظ أن سبب "ممارسة الرمي"، العامل الخاص في هذا المثال، ليس هو العامل العام، «القدرة الرياضية». بدلاً من ذلك، يساهم هذان العاملان بشكل مباشر في الاحترافية في رمي الكرة السريعة، ويوضح هذا الشكل رقم (٦-١٦). البنود A و B و C ناتجة مباشرة عن كل من "ممارسة الرمي" و "القدرة الرياضية". علاوة على ذلك، بعد الأخذ في الحسبان "القدرة الرياضية" العامة، فإن البنود التي تمثل "ممارسة الرمي" هي في الأساس أحادية البعد، وبالتالي يمكن أن تكون مناسبة للاندماج في مقياس. بشكل أساسي، فإن بنود "ممارسة الرمي" هي مجموعة أحادية البعد مندرجة ضمن النموذج الأوسع.

في حالات أخرى، سيكون النموذج الهرمي مناسباً لبياناتنا أفضل من النموذج ثنائي العوامل. لنفترض الآن أن مشكلة القياس التي نواجهها هي تحديد نموذج عاملي مناسب لمجموعة من مؤشرات وجهات النظر الاجتماعية والسياسية. قد تطلب البنود من المجيبين الإشارة إلى شعورهم تجاه سياسات الإنفاق الحكومي، مثل الدعم المالي لبرامج الفقر أو الشركات الصغيرة، وكيف ينظرون إلى القضايا الاجتماعية للاختيار الشخصي مثل زواج المثليين أو ملكية السلاح.



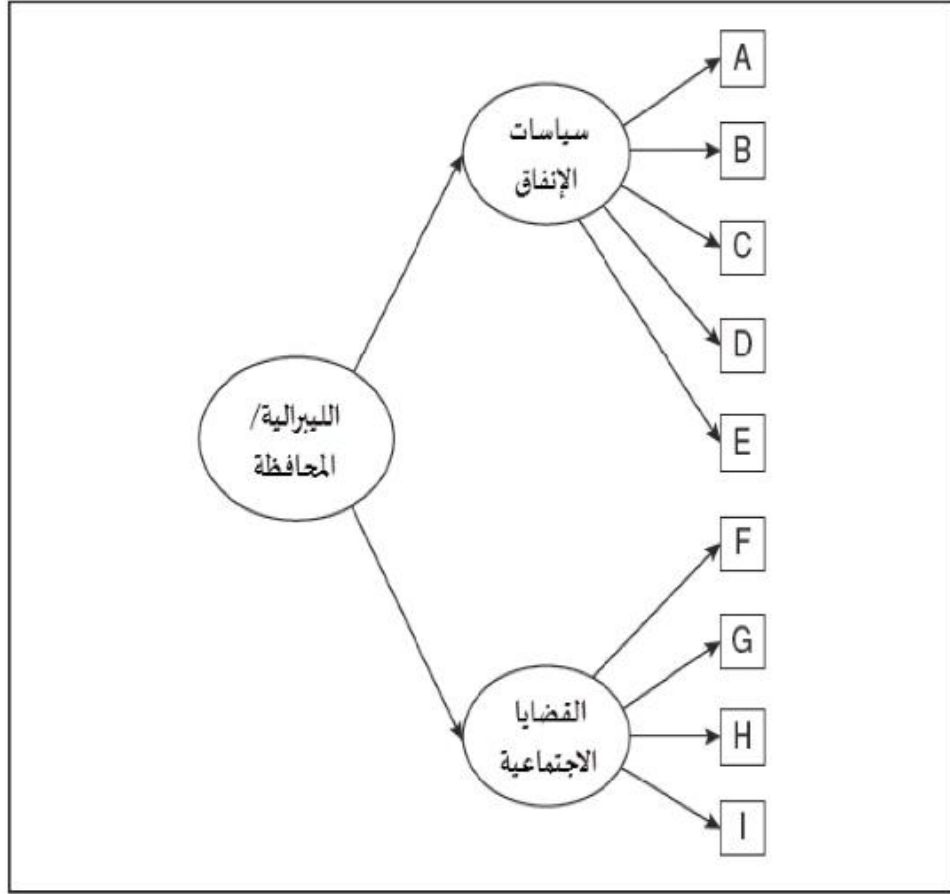
شكل (٦-١٦): نموذج ثنائي العوامل لـ "القدرة الرياضية"

كل فئة من فئات سياسة الإنفاق والقضايا الاجتماعية قد تحتوي على عدة بنود تمثل حالات خاصة لتلك المتغيرات. قد يكشف التحليل العاملي الأولي لهذه البنود المختلفة عن عاملين (من الرتبة الأولى) يشير محتواهما إلى أن أحدهما يستعرض وجهات النظر حول سياسات الإنفاق والآخر يستعرض المعتقدات حول الخيارات الشخصية. علاوة على ذلك، على الرغم من أن النتائج قد تشير إلى أن هذه العوامل متميزة بوضوح، إلا أنها قد ترتبط ببعضها البعض. قد يستنتج الباحث أنه بالإضافة إلى العاملين الأساسيين الخاصين، هناك عامل سببي أعلى رتبة مثل "الليبرالية / المحافظة" التي تحدد آراء المجيبين بشأن كل من سياسات الإنفاق والقضايا الاجتماعية. وبالتالي، فإن "الليبرالية / المحافظة" هي عامل محدد في وجهات النظر فيما يتعلق بكل من سياسات الإنفاق

والقضايا الاجتماعية، ووجهتها النظر المحكمة هما المحددان الفوريان للاستجابات للبنود الفردية. لاحظ أن "الليبرالية / المحافظة" ليست سبباً مباشراً للاستجابات للبنود. على سبيل المثال، قد يكون لبعض الأفراد الذين يعتبرون محافظين أو ليبراليين بشكل عام آراء حول القضايا الخاصة قيد النظر والتي تتعارض مع آراء معظم الليبراليين أو المحافظين. (مثال ملموس على ذلك هو أعضاء إدارة الرئيس الأمريكي ريجان المحافظة بشكل خاص، الذين لم يشاركوا آراء معظم المحافظين بعد محاولة اغتيال الرئيس فيما يتعلق بالسيطرة على الأسلحة). وعندما يكون الأمر كذلك، فإن العامل الثانوي البعد لن يكون لديه تأثير سببي كبير على الاستجابات لجميع البنود. أي إن التأثير السببي للعامل الثانوي على استجابات البنود يتم بواسطة العامل الأساسي بدلاً من أن يكون مباشراً. الترتيب المفاهيمي لهذه المتغيرات، مع تدفق السببية من عامل من الرتبة الثانية، إلى عامل من الرتبة الأولى، إلى البنود، يلائم نموذج تحليل العوامل الهرمي (انظر الشكل رقم ٦-١٧).

الغرض الأساسي من استخدام نهج النموذج ثنائي العوامل أو النموذج الهرمي لتحليل العوامل هو الكشف عن الأبعاد الكامنة المتعددة التي تمتلكها بعض البيانات وتحديد مجموعات البنود ذات البعد الأحادي بشكل أساسي. النموذج ثنائي العوامل مفيد بشكل خاص في هذا الصدد للعديد من أنواع المتغيرات التي تمثل مؤشرات مفهوماً بنائياً construct أكثر عمومية ومفاهيم بنائية متعددة أكثر خصوصية. يتيح نهج النموذج ثنائي العوامل للفاحص بشكل أساسي أن يضع العامل العام جانباً، ثم يدرس إلى أي مدى ينتج فعل ذلك مجموعات بنود أحادية البعد.





شكل (٦-١٧): النموذج الهرمي للآراء السياسية والاجتماعية

ومع ذلك، هذا الحل لا يخلو من المضاعفات والقيود. على سبيل المثال، ستظل الاستجابات المشاهدة للبنود تحتوي على تباين من كل من العوامل المشتركة والعوامل الخاصة. وبالتالي، من المرجح أن تؤدي التقييمات البسيطة للثبات التي لا تأخذ بالحسبان هيكلية النموذج ثنائي العوامل إلى المبالغة في تقدير الثبات. في الواقع، كما أشار ريس وآخرون (Reise et al., 2010)) أن كثيراً من الثبات الظاهر قد ينشأ عن العامل المشترك أكثر من العامل الخاص، حتى بالنسبة للبنود التي تشكل عاملاً خاصاً. وبالتالي، يجب اختيار تقديرات الثبات التي تعكس بدقة هيكل النموذج ثنائي العوامل للبيانات.

أيضاً، كما لاحظ ريس وآخرون (2010)، أن الحزم الإحصائية التقليدية للتحليل العاملي المعدة للنماذج بسيطة الهيكل تؤدي عملاً ضعيفاً في تحديد وجود عامل عام مصحوب بعوامل خاصة. وبالتالي، يجب على الباحثين استخدام برنامج مخصص صراحة لتحليل النموذج ثنائي العوامل (مثل الحزمة Psych ضمن بيئة تحليل البيانات R) لاستكشاف البيانات التي تناسب هذا النموذج (انظر Reise et al., 2010)، للحصول على مزيد من التفاصيل حول حزم البرامج المناسبة). يمكن أن تكون هذه التحليلات معقدة إلى حد ما وتستفيد من خبرة الباحث وحكمه. العديد من الباحثين غير المختصين في القياس النفسي قد يختار بحكمة عدم محاولة إجراء تحليلات ثنائية العوامل دون توجيهات إرشادية من الخبراء. ومع ذلك، من المفيد لجميع الباحثين المشاركين في القياس أن يكونوا على دراية بهذه الامتدادات من التحليل العاملي - وخاصة نموذج ثنائي العوامل - لأنها قد توفر بدائل مفيدة لنماذج الهيكل البسيط لتحقيق أحادية البعد الأساسية.

#### تفسير العوامل:

في المثال الذي يتضمن بنود «الوعي» و«الاعتمادية»، افترضنا أننا نعرف سلفاً بالضبط ما هي المتغيرات الكامنة. في كثير من الأحيان، ليس هذا هو الحال، وحينئذ سوف نعتمد على التحليل العاملي لإعطائنا دلالات بشأن طبيعة تلك المتغيرات الكامنة. ويتم ذلك عن طريق فحص البنود التي تمثل بقوة كل عامل (على سبيل المثال، البنود التي تحتوي على أكبر تشعبات على عامل معين). البنود ذات التشعبات الأكبر هي تلك الأكثر تشابهاً مع المتغير الكامن (وبالتالي، ترتبط بقوة أكبر). لذلك، يمكن أن توفر نافذة للتعرف على طبيعة العامل المعني. يتم القيام بذلك بسهولة عندما يكون هناك العديد من البنود التي تطرأ بوضوح على أحد المتغيرات العامة ذات التشعبات الكبيرة (على سبيل المثال، أكبر من 65.0) على نفس العامل. بالعودة إلى مثال الخصائص التي تعتبر مهمة لدى زميل العمل، إذا كانت الخصائص «ذكي» و«لديه ذاكرة حديدية» و«متعلم بشكل جيد»، وربما بند أو بندين آخرين ذات الصلة بالقدرة الفكرية تنتشع كلها بشكل كبير على نفس العامل، مع عدم وجود بنود لها تشعبات كبيرة على هذا العامل، فإنه سيكون من السهل إلى حد ما استنتاج أن «الأهمية المنسوبة إلى العقل» أو وصف ما مكافئ كانت تسمية ملائمة لهذا العامل.

على الرغم من أن اختيار تسمية لعامل ما قد يبدو واضحاً في بعض الحالات، فإن تعيين اسم لا يختلف عن إثبات الصدق validity. إذا كانت مجموعة البنود تستمر في الأداء كما يوحي الاسم

المعين، فإنه سيتحدد الصدق في النهاية. عندما تفسر العوامل تباينًا قليلًا نسبيًا ولديها بنود متعددة ومتباينة ظاهرياً وتتشعب على نحو مماثل، فإنه يجب أن يكون محلل العوامل حذرًا بشكل خاص في التفسير. إذا أسفر التحليل عن عامل واحد ببند تبدو مختلفة، فمن الأفضل عدم أخذ هذا العامل على محمل الجد كمؤشر على متغير كامن.

هناك نقطة أخرى تستحق أن نتذكرها في مرحلة التفسير وهي أن التحليل العاملي يمكن أن يعثر على البنية أو الهيكل التي تأخذ بالحسبان العلاقات فقط بين البنود التي تم تحليلها - لن تكشف بالضرورة عن طبيعة الظواهر في حد ذاتها. لا يمكن للباحث الذي يحاول تحديد الأبعاد الأساسية للشخصية، على سبيل المثال، الحصول على عامل “الانفتاح” إذا لم يتم تضمين أي بنود متعلقة بالانفتاح.

في بعض الأحيان، يمكن أن يؤدي تضمين عبارة أو جملة معينة إلى حدوث مظهر خاطئ لعامل ذي مغزى مفاهيمي. عندما تتم صياغة بعض العبارات بصيغة المتكلم (أنا، نحن)، والبعض الآخر بدون، على سبيل المثال، فإن ذلك قد يفسر نمط الارتباطات التي تمت مشاهدتها. كتوضيح، خذ بعين الاعتبار البنود الافتراضية التالية:

- 1- أنا أحب التفاح.
- 2- للبرتقال طعم جيد.
- 3- أنا أفضل التفاح على بعض الفواكه.
- 4- هناك العديد من الأشخاص الذين يحبون البرتقال.
- 5- أنا أستمتع بالتفاح بين الحين والآخر.
- 6- البرتقال عموماً رائحته لطيفة.
- 7- أنا أجد هشاشة التفاح جذابة.
- 8- يمكن أن يكون البرتقال الطازج حلوى لذيذة.

إذا كانت البنود الفردية يتم تشعبها على عامل واحد والبنود الزوجية على عامل ثانٍ، فلن نعرف ما إذا كانت الصياغة «أنا» للبنود الفردية هي سبب العاملين أو إذا كان الناس يعبرون عن مواقف متباينة تجاه النوعين من الفاكهة المذكورة. كل التفسيرات معقولة ولكن مربكة. هذه هي الحالة التي قد نقارن أو لا نقارن بها التفاح بالبرتقال.

### طريقة المكونات الرئيسية مقابل العوامل المشتركة:

هناك فئتان عريضتان من أساليب تحليل البيانات التي يعتبرها بعض المؤلفين أساساً نفس الشيء لكن آخرين يرون أنها مختلفة اختلافاً جذرياً، هما التحليل العاملّي وتحليل المكونات الرئيسية. أحياناً يستخدم مصطلح التحليل العاملّي لاحتضان كلتا التقنيتين وفي أحيان أخرى لوصف أحدهما على عكس الآخر. غالباً ما تستخدم مصطلحات العوامل المشتركة common factors والمكونات components كوسيلة أقل غموضاً للإشارة بشكل خاص إلى المركبات الناشئة عن التحليل العاملّي وتحليل المكونات الرئيسية، على التوالي. هناك أساس لتأكيد كل من التشابه والاختلاف بين هذه الأساليب.

تحليل المكونات الرئيسية تنتج واحداً أو أكثر من المتغيرات المركبة التي تلتقط الكثير من المعلومات الموجودة أصلاً في مجموعة أكبر من البنود. علاوة على ذلك، يتم تعريف المكونات على أنها مجاميع مرجحة للبنود الأصلية. أي إن المكونات الرئيسية هي تحويلات خطية للمتغيرات الأصلية. إنها تتجذر في البيانات الفعلية وتستمد من البنود الفعلية. إنها مجرد إعادة تنظيم للمعلومات في البنود الفعلية.

تحليل العوامل المشتركة ينتج أيضاً واحداً أو أكثر من المتغيرات المركبة التي تلتقط الكثير من المعلومات الموجودة أصلاً في مجموعة أكبر من البنود. ومع ذلك، تمثل هذه المركبات المتغيرات الافتراضية. ولأنها افتراضية، كل ما يمكننا الحصول عليه هو تقديرات لهذه المتغيرات. العامل المشترك هو مفهوم بنائي مثالي تخيلي الذي افتراضياً يسبب البنود المراد الإجابة عليها كما هي؛ ويتم الاستدلال على طبيعة المفهوم البنائي من خلال دراسة كيفية تأثيره على بنود معينة.

هل هما نفس الشيء أم مختلفان؟

يوضح الوصف أعلاه بعض الاختلافات بين المكونات والعوامل. أحد هذه الاختلافات هو أن العوامل تمثل متغيرات افتراضية مثالية نقدرها، بينما المكونات هي أشكال بديلة للبنود الأصلية من خلال ربط معلوماتها. الفكرة وراء استخراج العوامل المشتركة هي أنه يمكننا إزالة التباين من كل بند لا يشترك مع أي من البنود الأخرى. من منظور التحليل العاملي، كما كان الحال مع الثبات، فإن التباين غير المشترك هو خطأ في الأساس. وبالتالي، فإن التركيبات التي توصلنا إليها في استخراج العوامل المشتركة هي تقديرات للمتغيرات الافتراضية، الخالية من الأخطاء. وبهذا المعنى، تكون العوامل المشتركة مثالية - إنها تقديرات لما قد يبدو عليه متغير خالٍ من الأخطاء يحدد مجموعة من البنود. علاوة على ذلك، تحدد العوامل كيف تتم الإجابة على البنود، في حين يتم تعريف المكونات من خلال كيفية الإجابة على البنود. وبالتالي، في تحليل المكونات الرئيسية، تعتبر المكونات منتجات نهائية للبنود، وتحدد الدرجات الفعلية التي تم الحصول عليها من البنود طبيعة المكونات. بينما، في تحليل العوامل المشتركة فإننا نستحضر مفهوم المتغير الافتراضي المثالي الذي هو سبب درجات البند. العامل هو تقدير لهذا المتغير الافتراضي ويمثل سبباً، وليس أثراً، لدرجات البنود.

ماذا عن أوجه التشابه؟ هناك العديد:

أولاً، الفرق الحسابي بين الاثنين ضئيل. تذكر، في تحليل العوامل المشتركة، أن الهدف هو تقدير متغير مثالي خالٍ من الأخطاء. ولكن يجب علينا إنشاء هذا التقدير من البيانات الفعلية. كما أشرنا، تعتمد الطرق التحليلية للعوامل بشكل عام على مصفوفة الارتباط التي تمثل جميع الارتباطات بين البنود التي يجب أن تحلل إلى عوامل. في الفصل الثالث، أشرت إلى أن جميع القيم غير القطرية في مصفوفة التغاير أو الارتباط لا تمثل سوى التباين المشترك، الشيوع communal. كما أشرت آنذاك، أن مصفوفة الارتباط هي مجرد نسخة معيارية standardized من مصفوفة التباين - التغاير variance-covariance matrix. الارتباطات نفسها هي تغايرات معيارية، والوحدات unities هي تباينات معيارية للبنود. كل تباين معياري لبند يمثل كافة التباين، المشترك والخاص unique، الذي يظهره البند. لإنشاء متغير مثالي خالٍ من الأخطاء، يجب إزالة الجزء الخاص من التباين الموجود في تباينات البنود على طول القطر الرئيسي لمصفوفة الارتباط. وبشكل أكثر تحديداً، يجب استبدال كل وحدة unity بتقدير الشيوع communality، وهي قيمة أقل

من 0.1، تقارب فقط التباين المشترك لمتغير معين مع المتغيرات الأخرى المدرجة في التحليل العاملي. على سبيل المثال، إذا قدرنا أن هناك متغيراً معيناً يشارك 45٪ من إجمالي تباينه مع بنود أخرى في مصفوفة الارتباط، فسندد له تقديراً مشتركاً (الشيوع) بقيمة 45.0 ونضع تلك القيمة في المصفوفة، مع استبدال الرقم 0.1 الذي مثل التباين الكلي للبند. سنعمل ذلك لكل متغير، مع استبدال كل وحدة بتقدير الشيوع؛ (غالباً، يتم الحصول على تقديرات الشيوع عن طريق انحدار المتغير المعني على جميع المتغيرات المتبقية وذلك للحصول على مربع الارتباط المتعدد،  $R^2$ ، الذي يقوم بدور التقدير). تنتج عملية التصفية هذه مصفوفة ارتباط معدلة يتم استخدامها لاستخراج العوامل المشتركة بدلاً من المكونات، كما هو مبين في الجدول رقم (٦-١). التعويض بتقديرات الشيوع عن الوحدات  $unities$  هو الفرق الحسابي الوحيد الذي يميز استخراج العوامل المشتركة عن استخراج المكونات الرئيسية.

جدول (٦-١): مصفوفتا الارتباط لتحليل المكونات الرئيسية والتحليل العاملي المشترك

1.0	.70	.83	.48	.65
.70	1.0	.65	.33	.18
.83	.65	1.0	.26	.23
.48	.33	.26	1.0	.30
.65	.18	.23	.30	1.0

.45	.70	.83	.48	.65
.70	.52	.65	.33	.18
.83	.65	.62	.26	.23
.48	.33	.26	.48	.30
.65	.18	.23	.30	.58

ملاحظة: مصفوفة الارتباط في الأعلى، والتي تُستخدم لتحليل المكونات الرئيسية، تحتفظ بالوحدات في القطر الرئيسي. بينما تحتوي مصفوفة الارتباط في الأسفل، والتي تُستخدم للتحليل العاملي المشترك، على تقديرات الشيوع، بدلاً من الوحدة، على طول القطر الرئيسي.

ماذا عن قضية السبب والأثر؟ نحصل على كل من العوامل والمكونات من خلال تحليل الدرجات على البنود المشاهدة. كما يوضح شرح تقديرات الشيوخ، فإن العلاقات التجريبية بين البنود هي في النهاية أساس العوامل المشتركة. وهذا، بالطبع، صحيح أيضًا بالنسبة للمكونات. لذلك، حسابيًا، كلاهما يركز على البيانات التجريبية. علاوة على ذلك، فإن معظم محلي البيانات يصورون كلاً من المكونات والعوامل المشتركة كطرق لفهم المتغيرات الكامنة لمجموعة من البنود. وهذا يعني أن كلاً من المكونات والعوامل يُعتقد عادةً أنها تكشف سبب الدرجات المرصودة لمجموعة من البنود. في الواقع، يستخدم تحليل المكونات وتحليل العوامل، بالتبادل؛ حيث إنه في معظم الحالات التي تكون فيها البنود لديها شيء مشترك ذي معنى، فإن الطرق المختلفة تدعم نفس الاستنتاجات. في العديد من برمجيات تحليل العوامل (مثل PROCFACTOR في الحزمة الإحصائية SAS) يكون تحليل المكونات الرئيسية هو الافتراضي. بمعنى، يتم الاحتفاظ بالوحدات في مصفوفة الارتباط ما لم يتم تحديد الشيوخ. لذا، على الرغم من أن هناك أوجه تشابه واختلاف بين الاثنين، إلا أن التمييز بينهما غالبًا ما يتم تجاهله مع وجود عدد قليل إن وجد من الآثار السلبية.

ومع ذلك، فإن أحد الاختلافات المهمة، هو طبيعة التباين التي تفسرها المكونات مقابل العوامل. تأخذ المكونات في الحسبان الجزء المحدد من التباين الكلي بين المتغيرات الأصلية، في حين أن العوامل تأخذ في الاعتبار التباين المشترك أو الشائع. إن تقليل القيم القطرية لمصفوفة الارتباط، كما يفعل الشخص عند استخراج العوامل المشتركة، يقلل من البسط والمقام للنسبة التي تعبر عن نسبة التباين. لكنه يقلل من المقام إلى درجة أكبر بسبب الحسابات المحددة التي ينطوي عليها حساب الفروق ذات الصلة. ونتيجة لذلك، فإن نسبة التباين التي تفسرها مجموعة من المكونات والعوامل القابلة للمقارنة لن تكون متساوية أو مكافئة من الناحية النظرية. سوف تفسر العوامل نسبة أكبر من التباين الأكثر تقييدًا (أي التباين المشترك)، بينما تفسر المكونات نسبة أصغر من التباين الكلي. عند مناقشة النتائج التحليلية للعوامل وتقديم تقرير عن نسبة التباين المُفسَّر بالعوامل، فإنه من الأهمية بمكان أن نكون واضحين بشأن نوع التحليل (المكونات أم العوامل المشتركة)، وبالتالي، نوع التباين (المشترك أو الإجمالي) الذي يتم تفسيره.

هناك اختلاف آخر جدير بالملاحظة وهو أنه في بعض الحزم الإحصائية، فإن بعض المخرجات التي يتم الحصول عليها من استخراج العوامل المشتركة ستبدو بلا معنى. في كلا

النوعين من التحليل، سوف يتزايد المقدار التراكمي للتباين المفسر كلما تم استخراج العامل/المكون التالي. في العوامل المشتركة، غالبًا ما تتجاوز هذه النسبة 0.1 في مرحلة ما، وتستمر في الارتفاع مع إضافة العوامل التالية، وبعد ذلك، كما لو كان الأمر سحراً، تعود القيمة إلى 0.1 تماماً عند استخلاص العامل Kth. على الرغم من أن هذا يبدو غريباً، إلا أنه مجرد نتيجة ملفقة لطريقة حساب ويمكن تجاهلها. إذا استخدم محلل البيانات معايير معقولة لتحديد عدد العوامل التي يجب استخراجها، فإن العدد المختار سوف يسبق عادةً النقطة في تسلسل الاستخراج حيث تنشأ هذه الحالة الشاذة. ومع ذلك، فمن الممكن لهذا العدد المحدد من العوامل أن يفسر تقريباً كل (أي 100٪) التباين المشترك بين البنود الأصلية.

### التحليل العاملي التوكيدي Confirmatory Factor Analysis:

التفرع الآخر لأساليب تحليل العوامل يميز بين الطرق الاستكشافية والتوكيدية. تشير هذه المصطلحات في الأصل إلى نية محلل البيانات بدلاً من الطريقة الحسابية. وبالتالي، يمكن استخدام نفس التحليل على نفس المجموعة من البنود إما لتحديد بنيتها الأساسية (الاستكشافية) أو لتأكيد نمط معين من العلاقات المتوقعة على أساس نظرية أو نتائج تحليلية سابقة (توكيدية). مع التكرار المتزايد، تُستخدم هذه المصطلحات الآن للتمييز بين أنواع مختلفة من الأدوات التحليلية بدلاً من الأهداف البحثية المختلفة. عندما يستخدم الأشخاص مصطلح التحليل العاملي التوكيدي، فإنهم يتحدثون غالبًا عن طرق قائمة على نمذجة المعادلة البنائية structural equation modeling (SEM)). وعلى الرغم من أنه يجب استخدام هذه الأساليب بالطريقة التوكيدية بدلاً من الاستكشافية، إلا أنه يمكن استخدام الأساليب القياسية التحليلية للعوامل لأيٍ منهما. وبالتالي، فإن كلمة “التوكيدية” لا تعني بالضرورة نظام SEM.

ومع ذلك، فإن الأساليب المستندة إلى نمذجة المعادلات البنائية SEM تقدم بعض الفوائد الحقيقية التي تتميز بها على الأساليب التقليدية لتحليل العوامل في بعض الحالات. وتبرز هذه الفوائد لأن نماذج SEM مرنة للغاية. الشروط التي يتم افتراضها بواسطة أساليب تحليل العوامل التقليدية، مثل استقلالية حدود الخطأ للبنود عن بعضها البعض، يمكن تغييرها بشكل انتقائي في الأساليب المستندة إلى SEM. كذلك، فإن الطرق التقليدية تقيد محلل البيانات إما بالسماح للعوامل بالترابط مع بعضها البعض أو تتطلب أن تكون جميعها مستقلة عن بعضها البعض. الأساليب



القائمة على SEM يمكنها أن تجمع بين العوامل المترابطة وغير المترابطة إذا كانت النظرية تشير إلى أن مثل هذا النموذج ينطبق.

يمكن للطرق القائمة على نمذجة المعادلات البنائية SEM أيضاً توفير معيار إحصائي لتقييم مدى ملاءمة البيانات الحقيقية للنموذج المحدد. إن استخدامها بحكمة وتروي، يجعلها من الممكن أن تكون أصلاً. ولكن في بعض الأحيان، يمكن أن يؤدي ذلك إلى المبالغة في استخراج العوامل. حيث يؤدي استخراج المزيد من العوامل في كثير من الأحيان إلى تحسين ملاءمة النموذج. إن تطبيق معيار إحصائي صارم قد يحجب حقيقة أن بعض العوامل ذات الأهمية الإحصائية قد تمثل نسباً صغيرة غير مثيرة للاهتمام من التباين. خاصة في المراحل المبكرة من تطوير الأداة، قد يكون هذا مخالفاً لأهداف الباحث الذي يهتم بإيجاد أصغر عدد من العوامل المشبعة بالمعلومات بدلاً من حساب أكبر قدر ممكن من التباين.

وميزة أخرى تتوفر في الأساليب المعتمدة على نمذجة المعادلات البنائية SEM وهي الممارسة الشائعة لاختبار النماذج البديلة ومقارنة مدى ملاءمتها للبيانات. مرة أخرى، إن استخدامها بحكمة، يجعلها من الممكن أن تكون أداة قيمة. وعلى العكس من ذلك، إن استخدامها بإهمال، قد تؤدي إلى الوصول إلى نموذج ذي مواصفات أقل منطقية من الناحية النظرية ولكنها تؤدي إلى نموذج أفضل ملاءمة من الناحية الإحصائية. على سبيل المثال، قد تؤدي إزالة القيد المتمثل في عدم ارتباط أخطاء البنود مع بعضها البعض إلى ظهور قيم صغيرة جداً للارتباطات، ولكن قد لا يزال النموذج يتفوق من الناحية الإحصائية على نموذج مقيد. قد يقرر أحد الباحثين تجاهل الارتباطات الصغيرة لصالح نموذج أكثر بساطة، بينما يتم إقناع آخر بمعيار إحصائي لرفض البديل الأكثر بساطة. كمثال آخر، قد يكون النموذج الذي يفصل بين اثنين من العوامل المتميزة ولكن مترابطة للغاية (ربما مثل "الوعي" و"الاعتمادية") أفضل من النموذج الذي يربط بين الاثنين. إذا كان الارتباط بينهما عالياً جداً، فقد يبدو قرار إبقائهم منفصلين تعسفياً. على سبيل المثال، يمكن اعتبار الارتباط، لنقل 85.0، بين مؤشرين لنفس المفهوم البنائي construct دليلاً جيداً على تكافؤهما. لكن النموذج الذي يصف العوامل المنفصلة التي ترتبط مع بعضها البعض بمقدار 85.0 قد يناسب البيانات بشكل أفضل من النموذج الذي يجمع الاثنين في عامل واحد.

وليس المقصود من هذه التعليقات أن تشير إلى أن الطرق القائمة على نمذجة المعادلات البنائية SEM للتحليل العاملي التوكيدي هي سيئة. قدم ظهور هذه الأساليب مساهمات هائلة لفهم مجموعة متنوعة من قضايا القياس. ومع ذلك، أعتقد أن المرونة المتأصلة في هذه الأساليب تخلق فرصاً أكثر لاتخاذ قرارات رديئة، خاصةً عندما لا يكون محلل البيانات على دراية بهذه الأساليب. مع استثناء محتمل لتحليل المكونات الرئيسية (حيث تكون العوامل عبارة عن مركبات خطية في البنود)، لا توجد طريقة لتحليل العوامل تنتج حلاً صحيحاً فريداً. هذه الأساليب تنتج فقط حلولاً معقولة، والتي قد يكون هناك الكثير منها. ليس هناك ما يضمن أن النموذج الأكثر تعقيداً الذي يتفوق إحصائياً على نموذج أبسط بديل هو انعكاس أكثر دقة للواقع؛ قد يكون أو لا يكون. مع كل الأساليب التحليلية للعوامل، هناك حاجة إلى المنطق السليم لاتخاذ أفضل القرارات. التحليلات هي مجرد إرشادات لعملية صنع القرار وأدلة لدعم تلك القرارات. لا ينبغي لتلك التحليلات، في رأيي، أن تحل بالكامل محل صنع قرار الباحث. أيضاً، من المهم أن يتم وصف أساس القرارات، إحصائية أو غير ذلك، بدقة في التقارير المنشورة عن التحليل العاملي التوكيدي.

ملاحظة أخيرة حول هذا الموضوع: يعتبر الباحثون في بعض مجالات الاستقصاء (مثل أبحاث الشخصية) أن الحصول على نتائج متسقة من أساليب تحليل العوامل التقليدية كدليل تأكيدي أقوى من إظهار نموذج جيد ملائم وفقاً لمعايير إحصائية. على سبيل المثال، ذكر (Saucier and Goldberg 1996) أنه “نظراً لأن التحليل العاملي الاستكشافي يوفر اختباراً أكثر دقة للتكرار من التحليل العاملي التوكيدي، فقد يكون التحليل العاملي الاستكشافي غالباً مفضلاً على التحليل العاملي التوكيدي” (ص 35). السبب هو أنه إذا كانت البيانات المأخوذة من عينات مختلفة من الأفراد في مناسبات مختلفة قد أسفرت عن نتائج تحليلية للعوامل متطابقة بشكل أساسي باستخدام الأساليب الاستكشافية، فإن احتمال أن تكون هذه النتائج عبارة عن صدفة متكررة ضئيل للغاية. تذكر أنه في الأساليب المستندة إلى SEM لهذا الموقف نفسه، يحدد محلل البيانات العلاقات المتوقعة بين المتغيرات ويحدد برنامج الكمبيوتر ما إذا كان يمكن مطابقة هذا النموذج مع البيانات التجريبية empirical. بمعنى آخر، يُعطى الكمبيوتر تلميحات كبيرة حول الطريقة التي يجب أن تصير عليها الأمور. في المقابل، فإن إعادة اكتشاف هيكلية سابقة للعوامل دون اللجوء إلى مثل هذه التلميحات، كما قد يحدث مع التحليلات الاستكشافية المتكررة، يمكن أن تكون مقنعة للغاية.

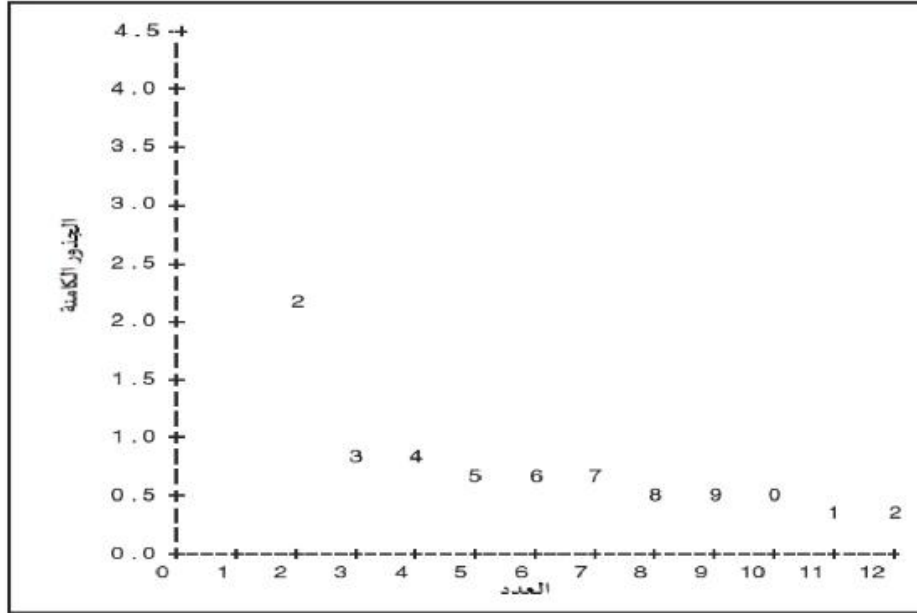
## استخدام التحليل العاملي في تطوير المقاييس:

يجب أن يجعل المثال التالي بعض المفاهيم التي تمت مناقشتها في هذا الفصل أكثر واقعية. طور بعض الزملاء وأنا (DeVellis et al., 1993) مقياساً لتقييم معتقدات الآباء بشأن من أو ماذا يؤثر على صحة أطفالهم. على الرغم من أن المقياس يحتوي على 30 بنداً ويقيم العديد من جوانب هذه المعتقدات، إلا أنه بالنسبة لهذا العرض التقديمي، سأناقش فقط 12 بنداً من تلك البنود:

- A- لدي القدرة على التأثير على رفاهية طفلي.
- B- ما إذا كان طفلي يتجنب الإصابة هو مجرد حظ.
- C- يلعب الحظ دوراً كبيراً في تحديد مدى صحة طفلي.
- D- يمكنني أن أفعل الكثير لحماية طفلي من الأذى.
- E- يمكنني أن أفعل الكثير لحماية طفلي من المرض.
- F- ما إذا كان طفلي يتجنب المرض هو مجرد مسألة حظ.
- G- الأشياء التي أقوم بها في المنزل مع طفلي جزء مهم من رفاهيته.
- H- سلامة طفلي تعتمد علي.
- I- يمكنني أن أفعل الكثير لمساعدة طفلي على البقاء معافى.
- J- الصحة الجيدة لطفلي هي إلى حدٍ كبير مسألة حسن حظ.
- K- يمكنني أن أفعل الكثير لمساعدة طفلي على أن يكون قوياً وصحياً.
- L- سواء كان طفلي بصحة جيدة أو مريضاً هي مجرد مسألة قسمة ونصيب.

تم تطبيق هذه البنود على 396 من الآباء والأمهات، وتم تحليل البيانات الناتجة باستخدام التحليل العاملي. كان الهدف الأول من تحليل العوامل هو تحديد عدد العوامل الكامنة وراء هذه البنود. وقد تم استخدام برنامج SAS لإجراء التحليل العاملي، وتنفيذ رسم الحصاة scree plot.

يُظهر رسم الحصاة - مشابهاً للنوع المطبوع بواسطة SAS - أدناه (انظر الشكل رقم ٦-١٨). لاحظ أنه تم رسم 12 عاملاً (أي ما يصل إلى عدد البنود)؛ ومع ذلك، يوجد 2 من هذه العوامل على الجزء الأولي من الرسم والباقي يشكل الحصاة التي تمتد على طول قاع الرسم. هذا يشير بقوة إلى أن 2 من العوامل مسؤولان عن الكثير من التباين بين البنود.



شكل (٦-١٨): رسم الحصاة الناتج عن التحليل العاملي للبنود المختارة

بعد تحديد عدد العوامل التي يجب الاحتفاظ بها، نقوم بإعادة تشغيل البرنامج الذي يحدد عاملين وتدويرهما بشكل متعامد (التباين الأعظم varimax). لو فشلنا في تقريب الهيكل البسيط، لقمنا بإجراء الدوران المائل oblique لتحسين الملاءمة بين البنود والعوامل. ومع ذلك، في هذه الحالة، أسفر الدوران المتعامد الأكثر بساطة عن مجموعات بنود ذات معنى لا لبس فيها وتشعبات قوية. وهذا يظهر جلياً في جدول تشعبات العوامل، كما هو موضح في الجدول رقم (٦-٢). يحتوي كل صف على تشعبات بند معين على العاملين. وقد قام أحد الخيارات المتاحة في برنامج SAS بإعادة ترتيب البنود الموجودة في الجدول بحيث يتم تجميع البنود ذات التشعبات العالية على كل عامل معاً.

في هذا الجدول، تم وضع خط تحت تشعبات البنود على العوامل التي تزيد عن 50.0. يتم تعريف كل عامل بالبنود التي يتم تشعبها عليه بشكل كبير (أي البنود التي تحتها خط). بالرجوع إلى محتوى تلك البنود، يمكن للمرء أن يميز طبيعة المتغير الكامن الذي يمثل كل عامل. في هذه

الحالة، فإن جميع البنود التي يتم تشبعها بقوة على العامل 1 تهم الوالد باعتباره عاملاً مؤثراً على ما إذا كان الطفل يبقى آمناً وصحياً.

جدول (٦-٢): تشبعات البنود على العاملين



بينما تلك البنود التي تنتشع في المقام الأول على العامل 2، تتعلق بتأثير الحظ أو القدر على صحة الطفل.

هاتان المجموعتان المتجانستان من البنود يمكن دراستهما بشكل أكثر. على سبيل المثال، يمكن حساب ألفا لكل مجموعة. حساب ألفا لمجموعات البنود هذه باستخدام برنامج SAS ينتج المعلومات الموضحة في الجدول رقم (٦-٣). كلا المقياسين يحتوي على معاملات ألفا للثبات بشكل مقبول. لاحظ أن الإجراء SAS CORR يحسب ألفا للبنود غير المعيارية والمعيارية. يكافئ حساب ألفا للبنود المعيارية استخدام صيغة ألفا المستندة إلى الارتباط. لكلا المقياسين، تسفر هاتان الطريقتان لحساب ألفا عن قيم متشابهة تمامًا. لاحظ أيضًا أنه لن تزيد قيمة ألفا بإسقاط أي من البنود لأي من المقياسين. قيم ألفا تقريبًا مرتفعة مثل تلك التي تم الحصول عليها للمقياس الكاملة عند إسقاط بند واحد (على سبيل المثال، البند H من مقياس 1 والبند B من مقياس 2). ومع ذلك،

فإن الاحتفاظ بهذه البنود يوفر بعض الضمان الإضافي بأن الثبات لن ينخفض عن المستويات المقبولة في عينة جديدة ولا تزيد أطوال المقاييس بشكل كبير.

جدول (٦-٣): معامل ألفا لجميع البنود ولجميع الـ  $1 - k$

مجموعات مركبة من البنود لمجموعتين مختلفتين من البنود

معامل ألفا كرونباخ للبنود غير المعيارية: 0.796472؛ وللبنود المعيارية: 0.802006				
البنود المعيارية		البنود غير المعيارية		
ألفا	الارتباط بالمجموع	ألفا	الارتباط بالمجموع	البند المحذوف
.749666	.676138	.741489	.675583	ITEM I
.755695	.644648	.748916	.646645	ITEM K
.775939	.535924	.770329	.545751	ITEM E
.769222	.572530	.763252	.562833	ITEM D
.787007	.474390	.782509	.466433	ITEM G
.799245	.404512	.793925	.409650	ITEM H
.793003	.440404	.785718	.437088	ITEM A
معامل ألفا كرونباخ للبنود غير المعيارية: 0.811162؛ وللبنود المعيارية: 0.811781				
البنود المعيارية		البنود غير المعيارية		
ألفا	الارتباط بالمجموع	ألفا	الارتباط بالمجموع	البند المحذوف
.749534	.682663	.748385	.684085	ITEM F
.776819	.594180	.775578	.596210	ITEM C
.763036	.639360	.762590	.636829	ITEM J
.777405	.592234	.776669	.593667	ITEM L
.806449	.493448	.806544	.491460	ITEM B



جميع المحاذير المطبقة على المقاييس بشكل عام في هذه المرحلة من تطويرها قابلة للتطبيق على المقاييس المشتقة بأساليب التحليل العاملي. على سبيل المثال، من المهم جداً تكرار ثبات المقاييس باستخدام عينة مستقلة. في الواقع، قد يكون من المفيد تكرار العملية التحليلية للعوامل بالكامل على عينة مستقلة لإثبات أن النتائج التي تم الحصول عليها لم تكن حدثت مرة واحدة مصادفة.

### حجم العينة:

إن احتمالية تكرار هيكلية العوامل هي جزئياً على الأقل دالة في حجم العينة المستخدمة في التحليل الأصلي. بشكل عام، سيكون نمط العوامل الذي ينبثق من التحليل العاملي للعينة الكبيرة أكثر ثباتاً من ذلك الناتج من عينة أصغر. حتماً، يطرح السؤال، “ما هو الحجم الكبير الكافي؟” هذا سؤال من الصعب الإجابة عليه (على سبيل المثال، MacCallum, Widaman, Zhang, & Hong, 1999). كما هو الحال مع العديد من الإجراءات الإحصائية الأخرى، ينبغي النظر إلى كل من نسبة عناصر subjects العينة لعدد المتغيرات التي يتم تحليلها والعدد المطلق لعناصر العينة، لكن ثمة عوامل أخرى مثل تشبعات البنود تلعب أيضاً دوراً مهماً في ذلك (MacCallum et al., 1999). كلما زاد عدد البنود التي يجب أخذها في الحسبان وكلما زاد عدد العوامل المتوقعة، ينبغي إدراج المزيد من العناصر في التحليل. من المغري، بناءً على هذه الحقيقة، البحث عن معدل قياسي لعدد العناصر إلى عدد البنود. ومع ذلك، كلما زاد حجم العينة تدريجياً، يمكن أن يتناقص معدل العناصر إلى البنود. بالنسبة لتحليل عاملي لـ 20 بنداً، من المحتمل أن تكون 100 عنصر قليلة جداً، ولكن بالنسبة لتحليل عاملي لـ 90 بنداً، قد يكون 400 عنصر مناسباً. يقترح كل من Tinsley and Tinsley (1987) معدلاً من حوالي 5 إلى 10 عناصر لكل بند وبما يصل إلى حوالي 300 عنصر. ويقترحان أنه عندما تكون العينة بحجم 300 عنصر، فإنه يمكن عدم التقيد بالنسبة بشكل صارم. في نفس الورقة، استشهدا بمجموعة أخرى من الإرشادات، منسوبة إلى Comrey (1973)، والتي تصنف عينة من 100 عنصر على أنها ضعيفة، و200 عنصر على أنها متوسطة، و300 عنصر على أنها جيدة، و500 على أنها جيدة جداً، و1000 على أنها ممتازة. ذكر كومري (1988) أن حجم العينة المكون من 200 عنصر مناسب في معظم حالات التحليل العاملي الاعتيادي التي لا تتضمن أكثر من 40 بنداً. على الرغم من أن علاقة حجم العينة بصلاحية

الحلول التحليلية للعوامل أكثر تعقيداً مما تشير إليه تلك القواعد القائمة على الخبرة، إلا أنه من الممكن أن تخدم الباحثين جيداً في معظم الظروف.

بالتأكيد ليس من غير المؤلف رؤية التحليلات العملية المستخدمة في تطوير المقياس استناداً إلى أحجام عينات أكثر تواضعاً (على سبيل المثال، 150 عنصراً). ومع ذلك، فإن النقطة الجيدة هي أن العينات الأكبر تزيد من تعميم الاستنتاجات التي تم التوصل إليها عن طريق التحليل العملي. بالطبع، قد يكون تكرار حل التحليل العملي على عينة منفصلة أفضل وسيلة لإثبات قابليتها للتعميم.

#### الخلاصة:

يعتبر التحليل العملي أداة أساسية في تطوير المقاييس. حيث يسمح لمحلل البيانات بتحديد عدد العوامل التي تقوم عليها مجموعة من البنود مما يمكن تنفيذ الإجراءات مثل حساب ألفا كرونباخ بشكل صحيح. بالإضافة إلى ذلك، يمكن أن توفر لنا رؤية ثاقبة لطبيعة المتغيرات الكامنة التي تقوم عليها مجموعة البنود.

#### تمارين:

- 1- اشرح مفهوم الهيكل البسيط ومدى ارتباطه باختيار العناصر لإدراجها في المقياس.
- 2- التحليل الموازي هو طريقة قائمة على الإحصاء لتحديد عدد العوامل أو المكونات الرئيسية التي يجب الاحتفاظ بها. صف بإيجاز كيف تحدد هذه الطريقة العوامل أو المكونات التي يجب الاحتفاظ بها.
- 3- افترض أن مجموعة من المتغيرات يتم تحليلها مبدئياً عن طريق استخراج العوامل المشتركة، وأن النتائج تكشف عن عامل أولي أساسي وعدة عوامل أصغر. علاوة على ذلك، يتم تشعب كل بند من البنود على كل من العامل الأول وواحد من العوامل الأصغر. يرغب الباحث في إنشاء مجموعة من المقاييس من هذه البنود، لكنه يشعر بالانزعاج من نمط التشعبات التي يكشف عنها هذا التحليل الأولي. ما الطريقة التحليلية للعوامل البديلة التي يمكن استخدامها لتحديد ما إذا كان يمكن تفسير البنود على أنها تنتمي إلى مجموعة من المجموعات أحادية البعد نسبياً؟

## الفصل السابع

### نظرة عامة على نظرية الاستجابة للبند (IRT)

نظرية الاستجابة للبند (IRT) هي بديل لنظرية القياس الكلاسيكية، وتسمى أيضاً نظرية الاختبار الكلاسيكية (CTT) Classical Test Theory. وقد لاقت هذه النظرية اهتماماً متزايداً في السنوات الأخيرة، وغالباً ما يتم تقديمها كبديل حديث ومتفوق على النظرية الكلاسيكية CTT (على سبيل المثال، De Boeck & Wilson, 2004, Embretson & Reise, 2010; Nering & Ostini, 2010; Reise & Revicki, 2015). كلتا النظريتان تشتركان في عدة خصائص أساسية. على سبيل المثال، أحادية البعد للمقياس تعد شرطاً لكل من نظريتي CTT وIRT. أي، يجب أن تشترك البنود في متغير واحد فقط إذا أريد دمجها في مقياس. بمعنى آخر، تحت هذين المنهجين، يجب أن تتقاسم البنود سبباً مشتركاً واحداً وبالتالي ترتبط ببعضها البعض. إذا كانت مجموعة من البنود متعددة الأبعاد (كما قد يكشف التحليل العاملي)، فإنه يجب التعامل مع مجموعات البنود أحادية البعد المنفصلة بشكل فردي. وهذا الأمر ينطبق في كل من النهج الكلاسيكي CCT ونهج نظرية الاستجابة للبند. لذلك، على سبيل المثال، إذا شكلت مجموعة مكونة من 50 بنداً خمسة عوامل منفصلة، فسيتم التعامل مع كل مجموعة من مجموعات البنود الخمسة بشكل منفصل، مما قد ينتج عنه خمسة مقاييس.

يُفرّق كلا الأسلوبين في القياس أيضاً بين تباين البند الذي ينشأ عن المستويات الفعلية للمتغير الكامن والتباين الذي ينشأ من الخطأ، لكن الطريقة التي يتم بها التمييز بين الدرجة الحقيقية true score والخطأ تختلف بين أسلوبَي القياس. الفكرة الأساسية الكامنة وراء نظرية القياس الكلاسيكية هي أن الدرجة المرصودة هي ببساطة نتيجة للدرجة الحقيقية بالإضافة إلى الخطأ. مصدر هذا

الخطأ لا تتم تجزئته إلى مصادر فرعية، مثل الاختلاف في الوقت أو البيئة أو البنود. بدلاً من ذلك، يتم جمع كل مصادر الخطأ في حد خطأ واحد.

بينما تميز طرق نظرية الاستجابة للبند الخطأ بشكل أكثر دقة، خاصة فيما يتعلق بخصائص البنود التي قد تؤثر على أدائها.

الهدف من نظرية الاستجابة للبند هو تمكين الباحث من تحديد خصائص معينة للبنود التي تكون مستقلة عن يجب عليها. وهذا مماثل للقياس المادي الذي يمكن فيه تقييم سمة كائن ما (مثل الطول أو الوزن) دون النظر إلى الطبيعة المحددة للكائن. عشرون رطلاً، على سبيل المثال، تعني نفس الشيء بغض النظر عما يتم وزنه. وبالتالي، يعطي مقياس الوزن التقليدي معلومات حول خاصية محددة للأشياء (أي الوزن) بغض النظر عن طبيعة الكائن الذي يتم وزنه. تطمح طريقة نظرية الاستجابة للبند أن تعمل الشيء نفسه مع بنود الاستبيان. الطرق الكلاسيكية تربط بعلاقة متأصلة بين أداة القياس والأشخاص الذين يتم قياسهم بينما لا تفعل طريقة نظرية الاستجابة للبند ذلك على الأقل من الناحية النظرية. على سبيل المثال، يتأثر ثبات المقياس وفقاً لنظرية القياس الكلاسيكية بالارتباطات بين البنود المكونة لهذا المقياس. فإذا كان الأفراد الذين تشكل بياناتهم أساساً لتقييم الثبات متماثلين للغاية في مستويات السمة الخاصة بهم، فإن مدى درجاتهم الحقيقية سيكون صغيراً. من نتائج قصر المدى أن تكون الارتباطات بين البنود أصغر، وبالتالي سيكون ثبات المقياس أقل بالنسبة للعينة التي يختلف فيها الأشخاص على نطاق أوسع فيما يتعلق بالسمة المقاسة. وبالتالي، فإن الثبات الذي يحصل عليه الفرد وفقاً لنظرية القياس الكلاسيكية لا تتعلق فقط بأداء أداة القياس، ولكن في بعض الظروف، تتعلق أيضاً بخصائص العينة قيد الدراسة. أساليب IRT لا تبني تقييمات الثبات على أساس بيانات العينة بالطريقة نفسها التي تتبعها أساليب القياس الكلاسيكية. بالطبع، في نهاية المطاف، لكلا الأسلوبين، يتم جمع معلومات البنود من الأفراد. وبالتالي، فإن تحقيق الميزة النظرية التي يمتلكها أسلوب نظرية الاستجابة للبند على أساليب القياس الكلاسيكية يعتمد على مدى تقييم البنود الموجودة تحت أسلوب نظرية الاستجابة للبند عبر عينات كبيرة غير متجانسة.

وصف جميع الاختلافات بين الأسلوب الكلاسيكي وأسلوب نظرية الاستجابة للبند خارج نطاق هذا الفصل. وللحصول على نظرة عامة عن أسلوب نظرية الاستجابة للبند، يمكن الاطلاع

على (Hambleton, Swaminathan, & Rogers (1991). بدلاً من محاولة إجراء مقارنة شاملة، في الأقسام التالية، سأركز على ثلاثة أوجه تمييز رئيسية بين أساليب القياس القائمة على نظرية القياس الكلاسيكية ونظرية الاستجابة للبند: (1) التركيز على البنود مقابل المقياس ككل، (2) تحديد البنود ذات المستويات المحددة للسمة التي يتم قياسها، و(3) التمثيل المرئي لخصائص البند والمقياس.

أول فرق مهم بين الأسلوبين هو أن نظرية الاستجابة للبند تولي اهتماماً كبيراً لخصائص البنود التي تشكل المقياس. في المقابل، تميل النظرية الكلاسيكية إلى التأكيد على خصائص المقياس ككل. كمثال على هذا الاختلاف، فكر في الثبات. عندما ناقشنا معامل ألفا كرونباخ في سياق القياس الكلاسيكي، لاحظنا أنه يمكن تعزيز ألفا إما عن طريق زيادة عدد البنود أو عن طريق تحسين متوسط الارتباط بين البنود. وبالتالي، فإن المزيد من البنود أو البنود الأفضل (بمعنى أن تكون مرتبطة بقوة أكبر مع المتغير الكامن) سوف يحسن الثبات. في نظرية القياس الكلاسيكية، غالباً ما يزداد ثبات المقياس من خلال التكرار - إضافة المزيد من البنود. عادةً، يتم تعزيز الثبات في طرق IRT ليس بالتكرار ولكن (حيثما أمكن ذلك) من خلال تحديد بنود أفضل. وهذا يعني أن نهج IRT ينظر إلى الثبات بشكل أساسي من منظور البنود، في حين تنتظر الأساليب الكلاسيكية إلى الثبات من منظور المقياس ككل. بينما يمكن للطرق الكلاسيكية، مثل فحص الارتباطات بين البنود والمجموع الكلي correlation total-item، تحديد كيفية مساهمة كل بند على حدة في الثبات الكلي للمقياس، إلا أن هذه العملية أقل ترسخاً في الأساليب الكلاسيكية منها في أساليب نظرية الاستجابة للبند لتطوير المقياس والتقييم.

يتمثل الاختلاف الثاني بين طريقتي القياس في أن نظرية الاستجابة للبند تفحص بشكل صريح مستوى السمة المقاسة الذي يؤثر بقوة على بند ما. قد يتم "ضبط" بنود مختلفة لمستويات مختلفة من السمة، وبالتالي، قد تكون حساسة للاختلافات على أجزاء مختلفة من السلسلة المتصلة الكاملة لتلك السمة. على سبيل المثال، ربما يقيس بندٌ مثل "أشعر أحياناً بالحزن" مستوى أقل من الحزن أو الاكتئاب مقارنة ببند مثل "أشعر أن الحياة لا تستحق العيش". قد يميز البند الأول شخصاً نادراً ما يشعر بمشاعر الحزن أو الاكتئاب عن الآخرين الذين يعانون من هذه المشاعر بشكل متكرر. وربما لا يكون مفيداً في التمييز بين الأشخاص الذين يشعرون بالحزن أحياناً عن أولئك

الذين يشعرون بالحزن في الغالب. من ناحية أخرى، قد يميز البند الثاني فقط بين الأشخاص في الطرف العلوي من السلسلة المتصلة للحزن أو الاكتئاب. وقد يفشل جميع الأشخاص على امتداد جزء كبير من هذه السلسلة في إقرار هذا البند، بينما الذين يعانون بشكل كبير من الحزن أو الاكتئاب يجيبون بشكل إيجابي. تحدد أساليب نظرية الاستجابة للبند هذين البندين على أنهما يمثلان نقاطاً مختلفة في سلسلة متصلة من الاكتئاب. من خلال تحديد البنود ذات الصلة بالنقاط على السلسلة المتصلة للمتغير الذي يتم قياسه، يمكن لطرائق نظرية الاستجابة للبند أن تساعد مطور الاختبار على تمثيل النطاق الكامل للسلسلة المتصلة للسمة التي يتم قياسها (الاكتئاب في هذا المثال) بالبنود المناسبة. بمعنى أنه يمكن تضمين البنود في مقياس يمثل مستويات منخفضة ومعتدلة وعالية من السمة بسهولة أكبر تحت نظرية الاستجابة للبند لأن هذه الطرق تربط البنود بشكل صريح بمستويات السمة التي تتعلق بها. في حين أن هناك طرق قياس كلاسيكية يمكن أن تحقق نتائج مماثلة (مثل فحص نسب المجيبين الذين يؤيدون كل بند وإجمالي الدرجات التي تصف المؤيدين وغير المؤيدين)، فإن هذه الطرق ليست خطوة عادية في طرق القياس الكلاسيكية. في المقابل، يعد تقييم «صعوبة» البند جزءاً لا يتجزأ من معظم طرق نظرية الاستجابة للبند. مدى قوة ارتباط بند ما بالمتغير الكامن وموقع البند على طول السلسلة المتصلة للسمة، يكون لهما تأثير على الثبات. البنود المرتبطة بقوة بالمتغير الكامن سترتبط أيضاً ارتباطاً قوياً مع بعضها البعض، وبالتالي زيادة متوسط الارتباط الداخلي بين البنود inter-item correlation وزيادة ثبات المقياس. علاوة على ذلك، من خلال تطوير بنود جيدة على طول النطاق الكامل للمتغير، يمكن للمرء أن يؤكد أن المقياس الناتج يمكن الاعتماد عليه عبر نطاق واسع من السمة التي يتم قياسها وليس فقط في أجزاء معينة منها. وبالتالي، يمكن أن يعزز كل من هذه الجوانب الشائعة والموجهة نحو البند في نظرية استجابة البند ثبات المقياس.

الاختلاف الثالث لنظرية الاستجابة للبند بالنسبة لطرق القياس الكلاسيكية يرتبط ارتباطاً وثيقاً بالأولين. تستخدم طرق نظرية الاستجابة للبند على نطاق واسع الأشكال الرسومية كوسيلة لتمثيل خصائص البنود والمقاييس ككل. سننظر إلى بعض هذه الصور قريباً، لكن أولاً، سأناقش أصول بعض المصطلحات المستخدمة عادةً لمناقشة خصائص البنود تحت نظرية IRT وترتبط ارتباطاً وثيقاً بالأشكال الرسومية التي تستخدمها نظرية الاستجابة للبند.

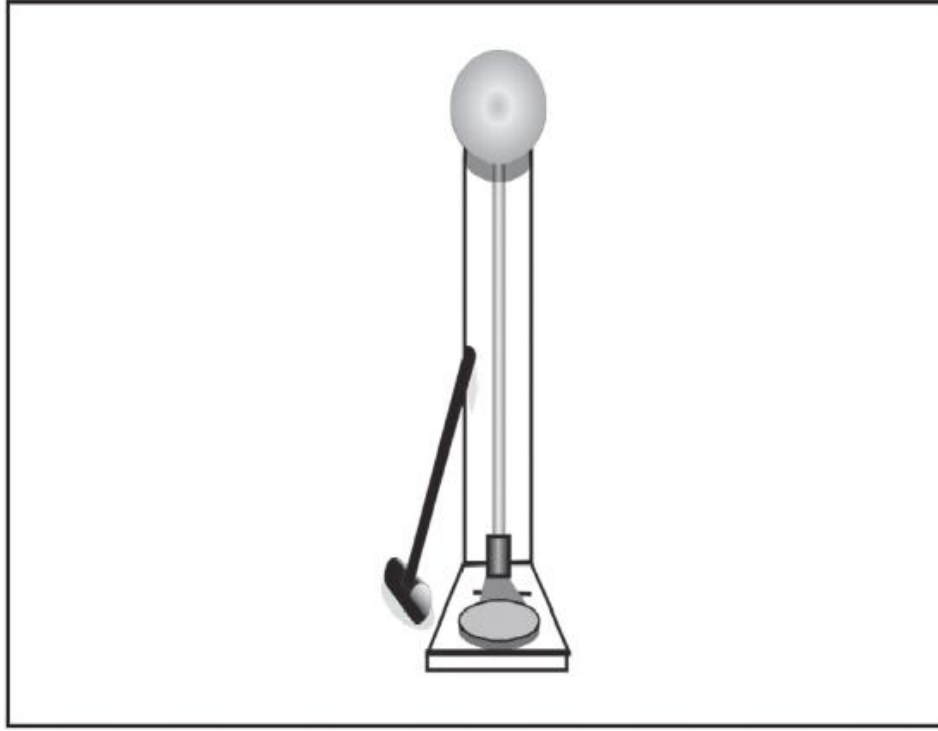
نظرًا لأن أسلوب نظرية الاستجابة للبند نشأ في سياق اختبار القدرات، فإن مفرداته vocabulary تحتوي على مصطلحات مرتبطة عادةً بمنطقة المحتوى تلك. أيضًا، نظرًا لتصنيف البنود في اختبارات القدرة على أنها “صحيحة” أو “غير صحيحة” (على الرغم من أن تنسيقها الأصلي قد يشتمل على أكثر من خيارين للاستجابة)، فإن التطبيقات التقليدية والأمثلة على نظرية الاستجابة للبند تتضمن بنوداً تأخذ واحدة من حالتين (على سبيل المثال، «الاجتياز» مقابل «الفشل»). من الأسهل مناقشة نظرية الاستجابة للبند مبدئيًا للبنود من هذا النوع، على الرغم من أنه لا يوجد سبب يمنع من توسيع الطرق المنبثقة عن النظرية (كما كانت بالفعل) لتشمل بنوداً ذات صيغ استجابة أخرى (مثل مقاييس Likert) تتطرق لمجالات المحتوى الأخرى. سننظر في البنود من هذا النوع الأخير في وقت لاحق من الفصل.

نظرية الاستجابة للبند هي في الحقيقة مجموعة من النماذج بدلاً من نظرية تحدد مجموعة واحدة من الإجراءات. إحدى الطرق المهمة التي تختلف بها النماذج البديلة لنظرية الاستجابة للبند هي عدد معلمات البنود التي هي محل الاهتمام. كان النهج المشترك في السنوات الأخيرة هو نموذج الثلاث معلمات three-parameter model، والذي، بشكل غير مفاجئ، يركز على ثلاثة جوانب من أداء البند. وهي صعوبة البند difficulty item وقدرته على التمييز item discrimination والتخمين guessing - أو بشكل أعم، قابليته للإجابيات الخاطئة susceptibility to false positives. أحد عناصر عائلة IRT في وقت مبكر، وما زال يتمتع بشعبية، هو نمذجة راش Rasch modeling (على سبيل المثال، Rasch, 1960; Wright, 1999)، والتي تحدد فقط معامل الصعوبة.

### صعوبة البند Item Difficulty:

على الرغم من أن مصطلح “صعوبة البند” أتى - كما يتضح - من اختبار القدرة، إلا أن المفاهيم التي يمثلها قابلة للتطبيق على نطاق أوسع. تشير صعوبة البند إلى مستوى السمة المقاسة المرتبطة بالانتقال من “الفشل” إلى “اجتياز” هذا البند. لقد شاهد معظمنا الأفلام القديمة التي تصور الكرنفالات أو المتنزهات التي تتميز ببعض العمل البطولي. “جهاز القياس” هو مسار عمودي ينتقل عبره الثقل. في الجزء العلوي من المسار يكون هناك جرس. في البداية، يقع الثقل في أسفل المسار عند نهاية لوح خشبي يمثل قاعدة الجهاز. يضرب “المشاركون” هذه القاعدة التي

تقع مقابل الثقل باستخدام مطرقة كبيرة، وبالتالي يتحرك الثقل إلى أعلى على طول المسار. هدفهم هو دفع الثقل بقوة كافية لضرب الجرس. لأغراضنا، يمكننا أن نفكر في الجهاز بأكمله باعتباره «البند» (انظر الشكل رقم ٧-١).



شكل (٧-١): جهاز افتراضي لاختبار القوة التي يؤدي فيها ضرب القاعدة بالمطرقة بقوة كافية إلى رنين الجرس

تتمثل صعوبة البند في مقدار القوة التي يجب أن يمتلكها «المستجيب» (أو، على نحو أدق، القوة التي يجب أن يبذلها المستجيب) من أجل «اجتياز» البند (على سبيل المثال، دق الجرس). بوضوح، يمكن للمرء بناء بنود مختلفة بدرجات مختلفة من الصعوبة (على سبيل المثال، «أشياء» أكثر صعوبة بمسارات أطول أو أثقال أكثر وزناً). ومع ذلك، يجب أن يكون من الممكن معايرة صعوبة جهاز معين يكون مستقلاً عن أي خاصية للشخص الذي يضرب بالمطرقة في تلك اللحظة.

نظرًا لأن هذا «الشيء» هو كائن مادي، فسيكون من السهل إلى حد ما تحديد مقدار القوة اللازمة للتسبب في رنين الجرس (مع التجاهل، في الوقت الحالي، تأثير ضرب القاعدة في مواقع مختلفة قليلاً عن نقطة الارتكاز). لذا، يُفترض أن يطلب مشغل الكرنفال جهازًا بحجم 10 أرطال أو



جهازًا يبلغ وزنه 100 رطل لتحقيق «معدل اجتياز» مرتفع أو منخفض حسب الأشخاص الذين يلعبون اللعبة. كل من تلك الأوزان قد يكون مناسبًا بشكل خاص لمجموعات مختلفة من العملاء (مثل الأطفال الذين يحضرون معرضًا للمدرسة مقابل البالغين الذين يحضرون معسكرًا رياضيًا).

يمكن للمرء أن يصف بنود استبيان بطريقة مماثلة. بالنظر، على سبيل المثال، إلى بنود قياس الاكتئاب، كما فعلنا في وقت سابق. يمكن للمرء أن يصمم البند ليكون «سهلاً نسبياً» (على سبيل المثال، «أشعر أحياناً بالحزن») أو «صعباً نسبياً» (على سبيل المثال، «أشعر أن الحياة لا تستحق العيش»). في الحالة الأولى، ستكون هناك حاجة فقط إلى قدر متواضع من السمة «الاكتئاب» «لاجتياز» (أي المصادقة) البند. لكن ألا يعتمد احتمال إصابة الشخص بهذا الشعور على من سئل؟ على سبيل المثال، إذا طرحنا هذا السؤال على الأشخاص المصابين سريريًا بالاكتئاب، فربما نجد أن نسبة أكبر من تلك العينة تؤيد هذا البند أكثر مما لو قمنا بطرح هذا السؤال على الجمهور العام. الهدف من تحديد صعوبة البند هو تحديد مقدار السمة المطلوبة لاجتياز البند بالمعنى المطلق. إذا كان ذلك ممكناً، فإن اجتياز الشخص لهذا البند له معنى ثابت فيما يتعلق بمستوى الاكتئاب، بصرف النظر عن كون هذا الشخص أو متوسط مستوى الاكتئاب في العينة قيد الدراسة. بمعنى آخر، يمكن وصف الشخص ليس فقط بالإشارة إلى عينة محددة ولكن من حيث المقياس المستقل عن أي عينة محددة.

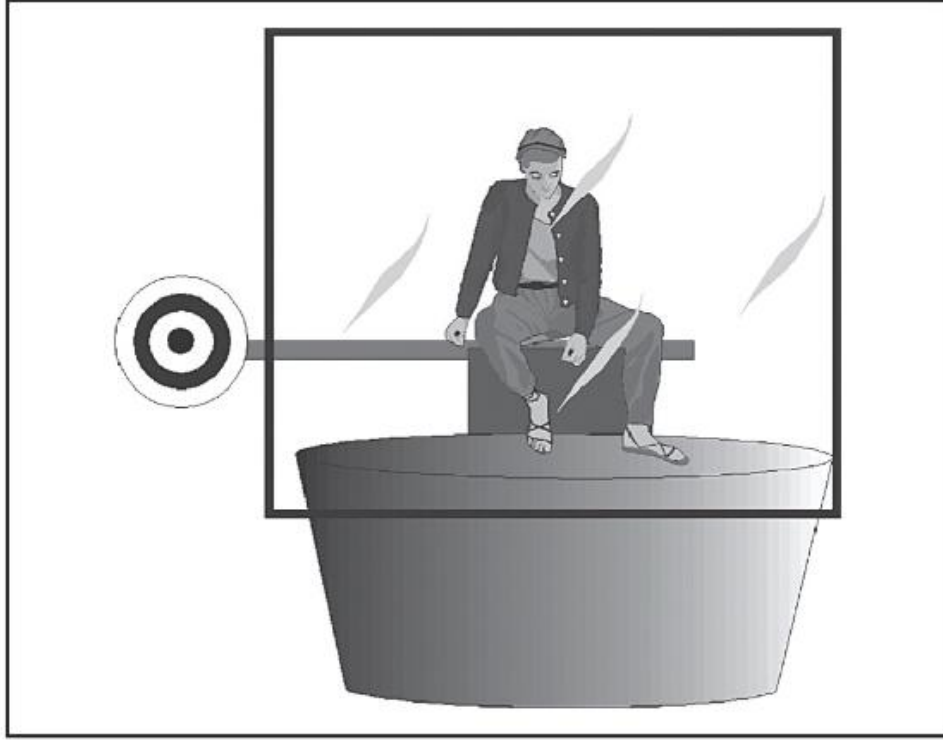
### تمييز البند Item Discrimination:

المعلمة الثانية التي تتناولها نظرية الاستجابة للبند IRT هي الدرجة التي تصنف بها الاستجابة كـ «اجتياز» أو «فشل» على أحد البنود بشكل لا لبس فيه. على نحو مختلف، كلما قل الغموض حول ما إذا كان الشخص «اجتاز» حقاً أو «فشل»، زاد التمييز في البند المعني. باستخدام تشبيه الجرس الكرنفالي الذي تمت الإشارة إليه سابقاً، قد تكون هناك مناسبات عندما يكون الثقل بالكاد يتصل بالجرس، مما يتسبب في اختلاف المراقبين حول ما إذا كان الجرس قد رن بالفعل أم لا. قد يسمع البعض رنيناً خافتاً، بينما لا يسمع الآخرون شيئاً على الإطلاق. ضمن نطاق القوة التي تدفع الثقل بحيث يلامس الجرس ولكنها لا تُنتج ما يتفق عليه الجميع على أنه صوت رنين واضح، فإن الجهاز في هذه الحالة يعطي معلومات غامضة. وبالنظر إلى هذا الغموض بطريقة أخرى، فإن نفس القوة المطبقة عدة مرات قد تؤدي إلى قيام المراقبين بتحديد أن الجرس قد رن في

بعض المناسبات ولكنه لم يرن في مناسبات أخرى. وبالتالي فإن قوة أكبر إلى حد ما سوف تنتج باستمرار رنيناً واضحاً لا لبس فيه، في حين أن قوة أضعف إلى حد ما سوف تنتج فشل (عدم رنين) لا لبس فيه على الدوام. ولكن هناك نطاق صغير من القوة التي يكون فيها الجهاز غامضاً. قد يعمل جهاز بديل بشكل مختلف وتنتج عنه نتائج أقل غموضاً. على سبيل المثال، قد تؤدي ملامسة الثقل للجرس إلى فصل التيار عن الدائرة الكهربائية ومن ثم تحريك قواطع التيار الكهربائي مما يتسبب في إضاءة المصباح ويظل مُضاءً حتى تتم إعادة ضبطه. إذا تم تصميم هذا الجهاز بشكل جيد، فمن المحتمل أن يسفر عن نتائج متسقة على نطاق صغير إلى حد ما من القوى، وبالتالي، سوف يميز بشكل أفضل من الجهاز القياسي. بدلاً من ذلك، فإن الجهاز الذي لم يكن له جرس على الإطلاق ولكنه بدلاً من ذلك طالب المراقبين برفع أيديهم إذا تجاوز الثقل خطاً محدد مسبقاً بجوار المسار الخاص به، من المحتمل أن ينتج عنه مزيد من الغموض، ومن ثم، فإن التمييز أقل. لذلك، فإن الجهاز أو البند الذي يميز جيداً يحتوي على جزء ضيق من نطاق ظاهرة الاهتمام التي تكون فيها النتائج غامضة. يحتوي الجهاز أو البند الأقل تمييزاً على مساحة أكبر من الغموض.

### التخمين، أو الإيجابيات الزائفة: Guessing, or False Positives

المعلمة الثالثة في نظرية الاستجابة للبند هي التخمين، أو الإيجابيات الزائفة. يعد مصطلح “التخمين”، مرة أخرى، إرثاً من أصول نظرية الاستجابة للبند في اختبار القدرة. في اختبار الاختيار من متعدد، يمكن للمستجيب الذي لم يكن يعرف الإجابة الصحيحة فعلياً اجتياز البند عن طريق تخمين خيار الجواب الصحيح. هذا التخمين هو إيجابي زائف. أي أنه يعطي إشارة إيجابية إلى أن المستجيب يمتلك مستوى معيناً من القدرة على الرغم من أنه لا يعرف بالفعل الإجابة الصحيحة على البند المعني. وبالتالي، فإن اعتبار المعلمة كإيجابية زائفة بدلاً من التخمين يسمح لنا بالتعميم بسهولة أكبر خارج سياق اختبار القدرة. لعل تشبيه كرنفال آخر قد يوضح فكرة إيجابي زائف. ربما تكون قد رأيت أكشاك يجلس فيها شخص خلف حاجز واقٍ من نوع ما فوق خزان مياه على منصة متصلة بذراع تمتد إلى جانب مرسوم عليه الهدف (انظر الشكل رقم ٧-٢).



شكل (٧-٢): جهاز افتراضي لقياس دقة الرمي يؤدي فيه ضرب الهدف بالكرة إلى انهيار المنصة وسقوط الشخص الذي يجلس عليها في خزان الماء

يُطلب من المتسابقين إلقاء كرات البيسبول على الهدف، مما يؤدي في حالة تعرضه لاصطدام الكرات إلى انهيار المنصة وسقوط الشخص في خزان المياه الذي يقع تحته. يمكننا التفكير في هذا الجهاز باعتباره «بنداً» يقيس دقة الرمي. يؤدي التسبب في سقوط الشخص من على المنصة في خزان المياه إلى «اجتياز» هذا البند. (الآن، يجب أن تكون قادرًا على وصف كيف يمكن أن تؤدي الاختلافات في الجهاز إلى زيادة أو تقليل صعوبة الجهاز وتمييزه). مع هذا الجهاز المحدد، يمكن للمرء أن يتخيل كيف يمكن أن تحدث «إيجابيات زائفة» - أي كيف يمكن للمشاركة بدون أي قدرة تقريبًا أن يسجل «اجتياز» عن طريق التسبب في سقوط الشخص الذي يجلس فوق الخزان. قد تكون إحدى الطرق هي أن «المشارك» يرمي بعنف ولكن الكرة تصيب الهدف فقط (يجب أن تذهب إلى مكان ما، بعد كل شيء). أو، بدلاً من ذلك، قد يتعطل الجهاز وقد تنهار المنصة تلقائيًا. في هذه الحالات، يقوم اللاعب / المشارك «بالاجتياز» ليس بناءً على القدرة ولكن بناءً على بعض الظروف غير ذات الصلة. وبالتالي، من الممكن «اجتياز» اختبار الرمي هذا حتى لو كانت قدرة الشخص قليلة أو معدومة. كما ذكر سابقًا، في سياق اختبار القدرات، تحدث

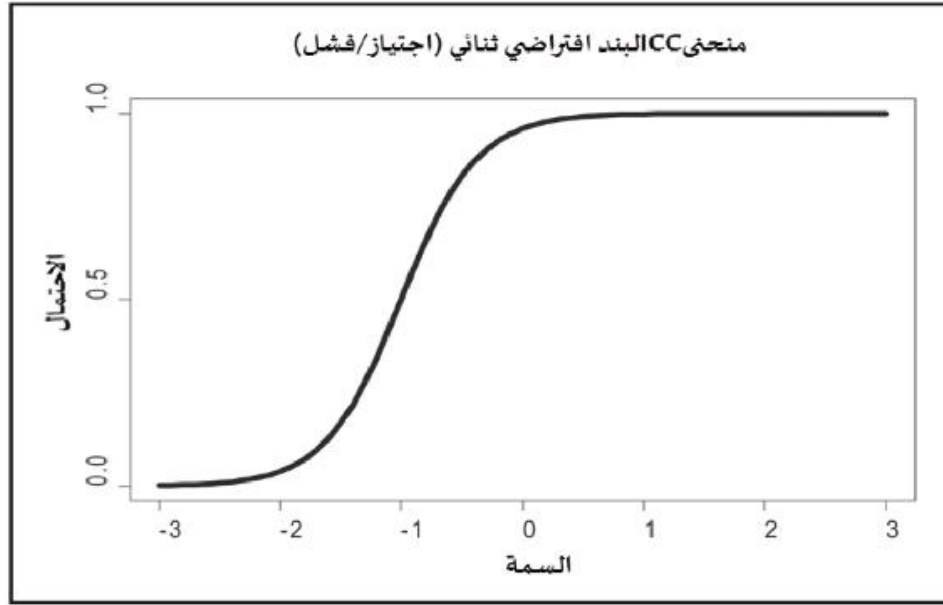
الإيجابيات الزائفة بشكل شائع نتيجة التخمين الناجح للإجابة الصحيحة على سؤال ما على الرغم من عدم معرفة الإجابة حقاً. (في سياقات القياس التي تكون فيها فرص التخمين أو الأنواع الأخرى من الإيجابيات الزائفة ضئيلة للغاية، مثل استخدام المقاييس لقياس الوزن، غالباً ما يكون نموذج المعلمتين - أي الصعوبة والتمييز - كافياً).

كل من معلمات البند الثلاثة هذه - الصعوبة والتمييز والإيجابيات الزائفة - لها علاقة واضحة إلى حدٍ ما بخطأ القياس. إذا كانت (أ) صعوبة بند ما غير مناسبة، أو (ب) تكون مساحة الغموض بين الاجتياز والفشل كبيرة، أو (ج) يشير البند إلى وجود سمة ما حتى في حالة غيابها، فإن البند يكون معرضاً للخطأ. تحدد نظرية الاستجابة للبند هذه الجوانب الثلاثة لأداء بند ما، وبالتالي توفر وسيلة لاختيار البنود التي من المحتمل أن تعمل بشكل جيد في سياق معين. الآن، لنرى كيف ترتبط هذه المعلمات الثلاث بالميزة الثالثة المهمة التي تميز أساليب نظرية الاستجابة للبند عن الأساليب الكلاسيكية، وذلك من خلال استخدام الرسوم البيانية لتوضيح خصائص البند.

### المنحنيات المميزة للبنود (ICC) ITEM-CHARACTERISTIC CURVES:

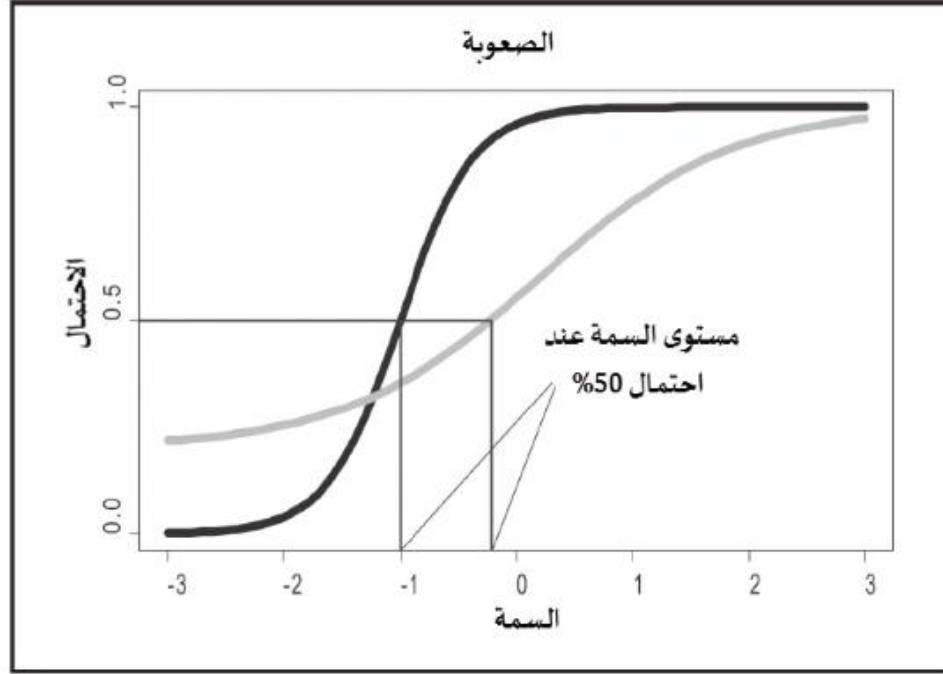
يمكن تلخيص صعوبة البند وتمييزه وتوليد إيجابيات زائفة في شكل المنحنى المميز للبند (ICC) الذي يمثل أداء البند بيانياً. عادةً، يكون منحنى ICC على شكل حرف S تقريباً، وتكشف أجزاء مختلفة من المنحنى عن معلومات حول كل من المعلمات الثلاثة محل الاهتمام.

يوضح الشكل رقم (٧-٣) ما قد يبدو عليه شكل المنحنى المميز للبند ICC. حيث يمثل المحور الأفقي X قوة الخاصية المميزة أو السمة التي يتم قياسها (مثل المعرفة أو القوة أو الدقة أو الاكتئاب أو الرغبة الاجتماعية أو أي ظاهرة أخرى قابلة للقياس). بينما يمثل المحور الرأسي Y احتمال "اجتياز" البند المعني بناءً على نسب درجات الفشل والاجتياز التي يتم ملاحظتها. إن معرفة كيفية استخدام المنحنى المميز للبند ICC لتقييم جودة البند يكون أسهل في الواقع إذا نظرنا إلى مخطط بياني يمثل بندين يمكننا مقارنتهما.



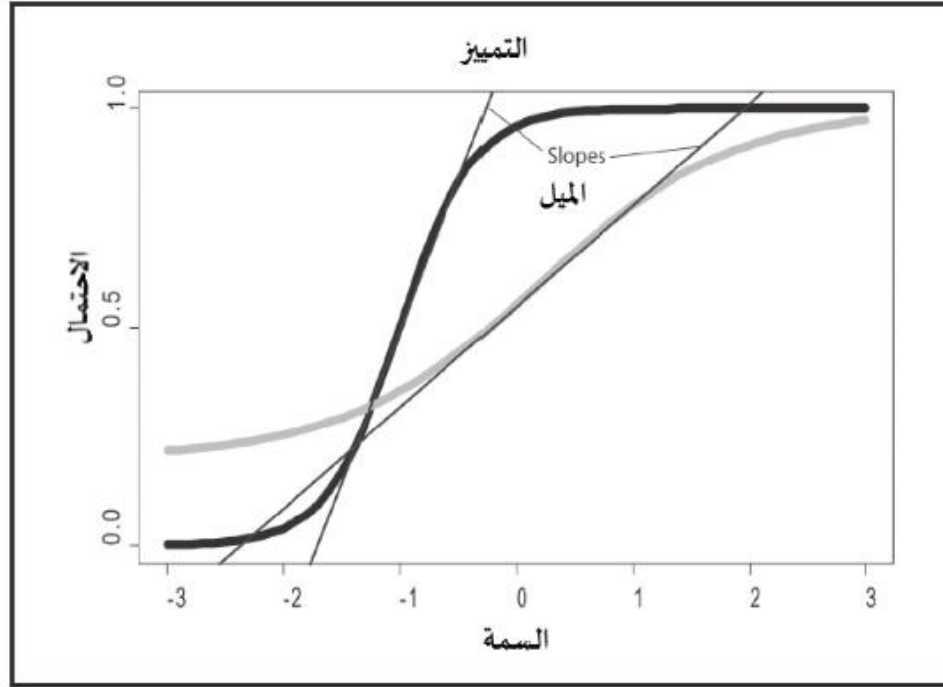
شكل (٧-٣): المنحنى الافتراضي المميز للبند (ICC) لبند ثنائي خيارات الإجابة (اجتياز/فشل)

يوضح الشكل رقم (٧-٤) صعوبة البند من خلال عرض منحنين. لاحظ أن النقاط التي يحقق فيها المنحنيان احتمالاً بنسبة 50% في اجتياز بنودهما مختلفة. بالنسبة للمنحنى الرمادي اللون، تكون هذه النقطة أبعد إلى اليمين. بمعنى، يجب أن يكون مقدار السمة أعلى للفرد حتى يكون لديه فرصة بنسبة 50% لاجتياز البند الذي يمثل المنحنى الرمادي مقارنةً بالبند الذي يمثل المنحنى أسود اللون. باستخدام هذا المعيار، يكون البند الذي يمثل المنحنى الرمادي أكثر صعوبة. الصعوبة في هذه الحالة ليست حكمًا ذاتيًا، وإنما وصفًا واقعيًا للنقطة على المحور الأفقي X المقابلة لتقاطع المنحنى لقيمة الاحتمال 50.0 على المحور الرأسي Y.



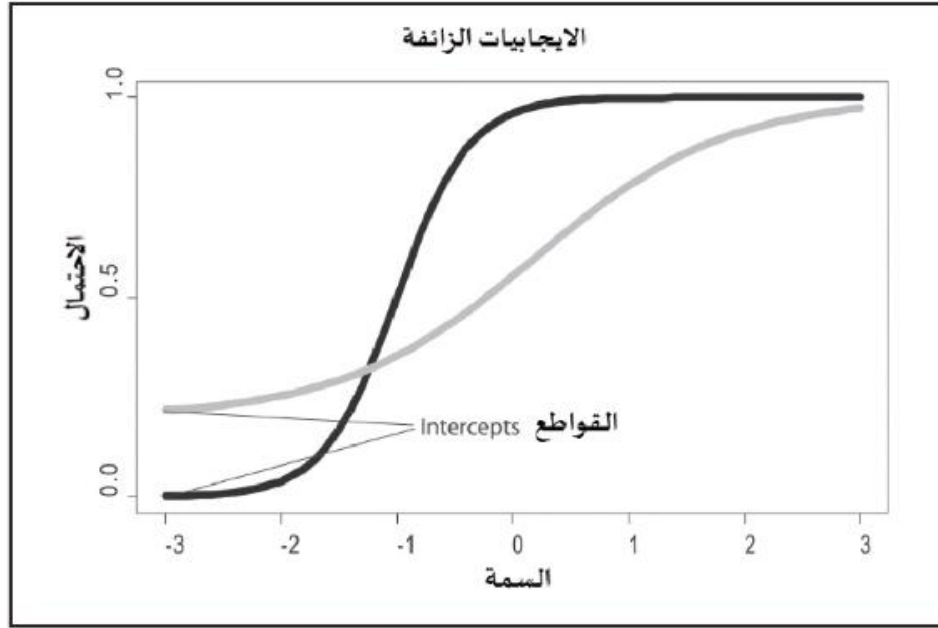
شكل (٧-٤): منحني ICC افتراضيان لبندين ثنائيين (اجتياز/فشل)  
مختلفين في درجة تحديد الصعوبة difficulty

يوضح الشكل رقم (٧-٥) كيف نقوم بتقييم التمييز discrimination باستخدام نفس المنحنيان المميزان للبند ICCs. يحتوي البند المقابل للمنحنى أسود اللون على ميل slope أكثر انحدارًا عند نقطة الاجتياز 50% من البند الذي يمثل المنحنى رمادي اللون. والنتيجة هي أن زيادة أقل في السمة تعطي زيادة أكبر في الثقة بأن المستجيب سيجتاز هذا البند مما هو عليه بالنسبة للبند الذي يمثل المنحنى الرمادي اللون. لذلك، يكشف المنحنى أسود اللون أن منطقة المحور الأفقي X التي تتوافق مع درجة غامضة أصغر من المنطقة المكافئة للبند الآخر. وبالتالي، فإن بند المنحنى أسود اللون يميز بين أولئك الذين يفشلون والذين يجتازون بشكل أكثر فعالية من بند المنحنى رمادي اللون.



شكل (٥-٧): منحني ICC افتراضيان لبندين ثنائيين (اجتياز/فشل)  
مختلفين في تحديد درجة التمييز **discrimination**

أخيراً، في الشكل رقم (٦-٧)، يمكننا أن ننظر إلى ميول البندين لإعطاء درجات الاجتياز حتى عندما تكون قدرة (أو أيًا كانت السمة التي يتم قياسها) المستجيب في الأساس تساوي صفراً. وكما كنت أنت قد خمنت، يتم تحديد ذلك من خلال النقطة التي يتقاطع فيها المنحنى المميز للبند ICC مع المحور الرأسي  $Y$ . بالنسبة للبند الممثل بالمنحنى أسود اللون، تكون قيمة القاطع  $intercept$  صفراً. وبالتالي، فإن احتمال قيام شخص ما باجتياز البند إذا كان يفتقر تمامًا إلى السمة المعنية سيكون صغيراً جداً. بالنسبة للبند الممثل بالمنحنى رمادي اللون، فإن هناك احتمالاً كبيراً (حوالي 20٪) بأن شخصاً لا يتمتع بالسمة أو القدرة المقاسة سيجتاز البند، وبالتالي، لا يمكن تمييزه، استناداً إلى البند المعني، من شخص ذي قدرة عالية. (لاحظ أن هذا سيكون هو الحال مع بند متعدد الخيارات له خمس إجابات محتملة، إحداها كانت صحيحة). يوضح الرسم التخطيطي في الشكل رقم (٦-٧) الاختلافات في قواطع  $intercepts$  المحور الرأسي  $Y$  للبندين - التي تعد الأساس لاستنتاج ذلك. مرة أخرى، البند الممثل بالمنحنى أسود اللون هو الأفضل أداءً.



شكل (٦-٧): منحني ICC افتراضيان لبنتين ثنائيتين (اجتياز/فشل)  
مختلفين في درجة تحديد الإيجابيات الزائفة false positives

من الناحية النظرية، يمكن استخدام نظرية الاستجابة للبند IRT لتحديد المَعْلَمَات لكل بند من البنود. بعد ذلك، اعتماداً على تفاصيل التطبيق المراد، يمكن استخدام البنود ذات الأداء الأمثل لمعالجة مشكلة القياس الموجودة. على سبيل المثال، يمكن أخذ عينة من البنود “السهلة” لتطبيقها على الأشخاص الذين لديهم مستويات منخفضة نسبياً من القدرة (السمة) المعنية وأخذ عينة من البنود الصعبة لأولئك الذين لديهم مستويات عالية من القدرة أو السمة المعنية. هذا يشبه بشكل مباشر استخدام جهاز جرس يزن 10 أرطال في معرض أو كرنفال مخصص للأطفال وجهاز جرس يزن 100 رطل في معسكر للرياضيين البالغين. يمكن أن يؤدي استخدام البند الخطأ - مثل استخدام جهاز الجرس الخاطئ - إما إلى الإحباط (إذا كانت المهمة صعبة للغاية) أو نقص الدافع (إذا كانت المهمة سهلة للغاية). أيضاً، إذا كان المقياس الذي سيتم توليفه سيكون أساساً لقرارات مهمة، فإن تقليل نطاق الغموض لكل بند واحتمال ظهور إيجابيات زائفة هي أيضاً احتمالات جذابة.

استخدام الاختبار التكيفي adaptive testing هو امتداد لهذه الفكرة. في الاختبار التكيفي، البنود المطبقة على كل فرد يمكن تصميمها وفقاً لمستوى السمة عند ذلك الفرد. إن اجتياز بند سابق أو الفشل في ذلك، يمكن أن يرشد إلى اختيار البند التالي، مع إعطاء الأفضلية لبند تم ضبطه على



مستوى القدرة الذي تشير إليه الإجابات السابقة. تتم إدارة هذه العملية عادةً بواسطة جهاز كمبيوتر وغالبًا ما يشار إليها باسم الاختبار التكيفي المحوسب computerized adaptive testing أو CAT. على الرغم من أن هذه المنهجية تتطلب مجموعة كبيرة من البنود التي تمثل مجموعة واسعة من مستويات القدرة، إلا أنها أداة قوية تتيحها طرق نظرية الاستجابة للبند IRT.

تتمتع طرق نظرية الاستجابة للبند IRT بميزة واضحة تتمثل في توجيه انتباهنا إلى ثلاثة جوانب (نسخة المعلومات الثلاثة الشائعة حاليًا) لأداء بند ما. من خلال الأساليب المتجذرة في نظرية القياس الكلاسيكية، قد نعرف (على سبيل المثال، من أدائها في التحليل العائلي أو حسابات معامل ألفا) ما إذا كان بند ما يعمل بشكل جيد أو ضعيف ولكن قد لا يكون لدينا فهم واضح لطبيعة أي أوجه قصور لديه. بالمقابل، قد تساعدنا طرق نظرية الاستجابة للبند IRT على تقييم نقاط القوة والضعف في بند ما بشكل أكثر تحديدًا.

#### تطبيق نظرية الاستجابة للبند IRT على البنود متعددة الاستجابة:

تنطبق مقدمة نظرية الاستجابة للبند المقدمة أعلاه على نماذج ثلاثية المعلومات وتتعامل مع الاستجابات ثنائية التفرع، مثل صح أو خطأ. كما اقترح سابقًا، هناك العديد من نماذج نظرية الاستجابة للبند. وإنه من الشائع في العلوم الاجتماعية والسلوكية، أننا نتعامل مع أنواع من خيارات الاستجابات المرتبة ومتعددة المستويات والتي تمت مناقشتها بوجه عام في الفصول السابقة.

بالنسبة للأدوات ذات خيارات الاستجابة متعددة المستويات، مثل مقاييس ليكرت Likert، يتم تطبيق نماذج نظرية الاستجابة للبند IRT المكيفة بشكل خاص لتلك الأدوات. أحد أكثر هذه النماذج شيوعًا هو نموذج الاستجابة المتدرج model response graded الذي طورته Fumiko Samejima (على سبيل المثال، Samejima, 1969) ويطبق بشكل شائع عن طريق برنامج s Multilog software'David Thissen' (Thissen Chen & Bock, 2003) يوفر هذا النموذج معلومات حول كيفية ارتباط كل خيار من خيارات الاستجابة المتعددة بالقدرة. فبالنسبة للبند الجيد، يجب أن يشغل كل خيار استجابة جزءًا متميزًا أو أقل من القدرة المقاسة.

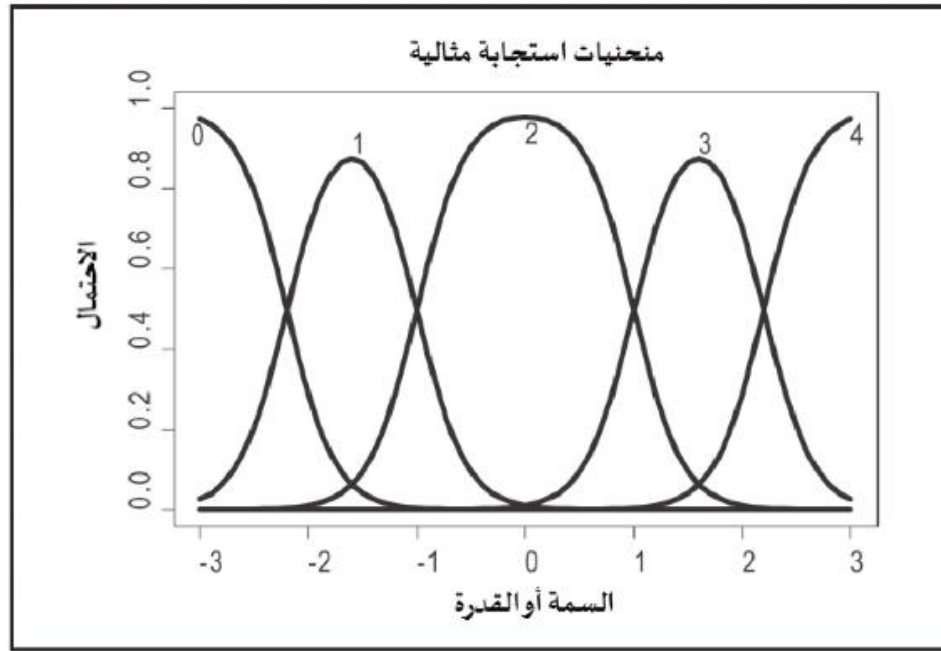
افترض، على سبيل المثال، بنداً مثل "أشعر بالدوار عندما أستيظ للمرة الأولى في الصباح" وله خيارات الاستجابة التالية: (0) "أبداً"، (1) "نادرًا"، (2) "بعض الوقت"، (3)

“معظم الوقت” و(4) “دائمًا تقريبًا”. يتوقع المرء أن احتمال اختيار خيارات أعلى تدريجيًا على المقياس يرتبط بمستويات أعلى تدريجيًا من الدوار (الدوخة) الصباحي (أي مستويات أعلى تدريجيًا من السمة محل التقييم). بطريقة ما، يحل بند مثل هذا محل العديد من البنود ثنائية الاستجابة. لذا، فبدلاً من البنود الثنائية من النوع “نعم - لا” المنفصلة، مثل “لا أشعر بالدوار أبدًا عندما أستيظ للمرة الأولى في الصباح”، “نادرًا ما أشعر بالدوار عندما أستيظ في الصباح”، “في بعض الأحيان أعاني من الدوار عندما أستيظ في الصباح”. وهكذا، يتم تقديم بند واحد له العديد من خيارات الاستجابة التي يمكن للشخص المستجيب الاختيار من بينها.

إذا كان البند يعمل بشكل جيد على مجتمع معين، فإن احتمالية اختيار خيار الاستجابة الأول (أي “أبدًا”) يجب أن تكون مرتفعة للأشخاص ذوي المستويات المنخفضة للسمة (الدوخة) ويجب أن تتضاءل فعلياً إلى الصفر مع الزيادة في مستوى السمة. لذلك، ينبغي أن يكون المنحنى الذي يمثل احتمال اختيار خيار الاستجابة الأول عالياً في أقصى اليسار (المنطقة المقابلة لمستويات منخفضة من الدوخة في الصباح) ومنخفض عند أقصى اليمين. في الطرف الآخر من المقياس، يجب أن يكون احتمال اختيار خيار الاستجابة الأخير (أي “دائمًا تقريبًا”) صفرًا تقريبًا للأشخاص في المستويات المنخفضة إلى المتوسطة من السمة، مع التقدم والاقتراب من نسبة 100٪ من احتمالية الدوخة للأشخاص الذين يعانون من أعلى مستويات الدوخة في الصباح. يجب أن يكون منحنى هذه الاستجابة منخفضًا للغاية عند النهايات المنخفضة من مستويات الدوخة الصباحية، بحيث يصل إلى الحد الأقصى للقيمة عند أقصى اليمين من مستويات الدوخة. وبالتالي، فإن منحنى كلٍ من هذين الخيارين للاستجابة القصوى غير متناظرة nonsymmetrical، حيث إن أحد الطرفين منخفض للغاية والآخر مرتفع جداً. وأيضاً، يجب وضع كلٍ من هذين المنحنيين على أقصى حد من النطاق المطابق لمستوى السمة التي يمثلها خيار الاستجابة. على النقيض من ذلك، يجب أن تكون المنحنيات التي تقع بين خيارات الاستجابة متماثلة إلى حدٍ ما، مع احتمال منخفض لاختيار المستجيبين للخيارات الواقعة عند أطراف السمة واحتمال عالٍ عند نقطة ما على النطاق المناسب لخيار الاستجابة للبند المعني. بالنسبة لخيارات الاستجابة الأعلى، يجب أن تظهر القمة عند مستويات أعلى من نطاق السمة مقارنةً بخيارات الاستجابة الأقل.

يمكننا تصوير احتمالية اختيار كل خيار من خيارات الاستجابة، بناءً على مستوى السمة التي يتم قياسها لدى المستجيب، من خلال سلسلة من منحنيات الاستجابة للفئة -category-response curves. قد تبدو مجموعة كاملة من منحنيات الاستجابة للفئة الافتراضية لخمس خيارات استجابة مستخدمة في المثال السابق بشكل يشبه ما هو موضح في الشكل رقم (٧-٧).

في الرسم التوضيحي المثالي الموضح في الشكل رقم (٧-٧)، يتوافق كل منحنى مُعَلَّم برقم مع خيار الاستجابة في مثال الدوخة الصباحية الموضح سابقًا. يمثل خط الأساس قوة السمة التي يتم قياسها من خلال البند. ويتركز مقياس القدرة عند الصفر ويمكن تفسيره بطريقة تشبه الدرجات المعيارية.

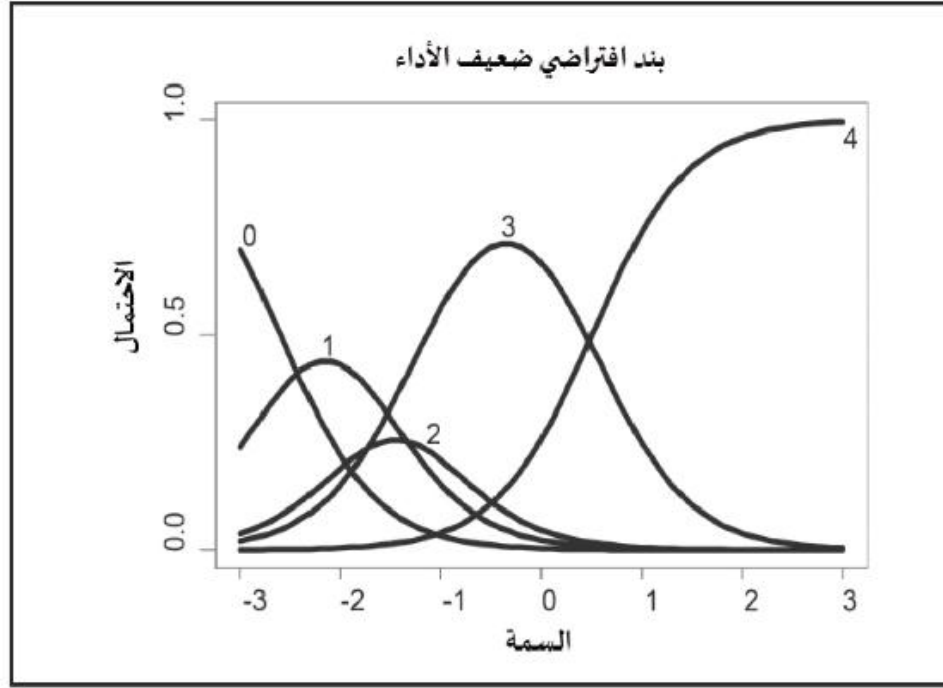


شكل (٧-٧): منحنيات استجابة الفئة المثالية لخمس خيارات استجابة

لاحظ أن هذا الشكل التوضيحي هو مثال مثالي وليس شكلاً بيانياً حقيقياً مبني على بيانات فعلية. يحتفظ هذا المثال بخاصية منحنيات استجابة البند الحقيقية: بالنسبة لأي مستوى من السمة، فإن مجموع المنحنيات المختلفة عند تلك النقطة على خط الأساس يساوي واحد. لذلك، لأي خط رأسي مرسوم عند أي نقطة على مقياس السمة، فإن القيم (من المحاور الواقع على الجهة اليسرى) للنقاط التي يعبر بها ذلك الخط منحنيات استجابة الفئة سيكون مجموعها يساوي الواحد الصحيح.

بمعنى آخر، فإن احتمال اختيار استجابة ما يساوي الواحد الصحيح لأي مجيب بمستوى من السمة محل القياس.

يوضح الشكل رقم (٧-٨) مثلاً أقل مثالية ولكنه لا يزال افتراضياً. يشير موقع منحنى كل خيار من خيارات الاستجابة الواقعة على امتداد بُعد السمة إلى “صعوبة” ذلك الخيار. ويمكننا أن نرى أن القمة لكل منحنى من منحنيات خيار الاستجابة المتتالية مرتبة كما كنا نتوقع، وتقع خيارات الاستجابة التي تشير إلى دوخة صباحية أقل على يسار خيارات الاستجابة التي تشير لدوخة صباحية أكثر. وهذا يعني أن الإشارة إلى المزيد من الدوار مرتبط بمنحنى يقع إلى الأعلى (أي إلى أقصى اليمين) على امتداد نطاق السمة، كما كنا نأمل. خاصية أخرى ملحوظة في الشكل رقم (٧-٨) هي أن منحنيات الاستجابة تبدو مزدحمة نحو الجانب الأيسر من المقياس. وهذا يعني أن السؤال يميز بشكل أكثر دقة بين الأشخاص الذين يعانون من دوخة قليلة نسبياً مقارنةً بالأشخاص الذين يعانون من الدوار كثيراً. إذا أخذنا بعين الاعتبار النطاق على مقياس السمة الواقع بين (-٥,٥ و ١,٥)، فإننا نرى أن جميع خيارات الاستجابة الخمسة يتم اختيارها أحياناً من قبل المستجيبين الذين لديهم ذلك النطاق من قيم الدوخة، على الرغم من أن خيارات الاستجابة 1 و 3 يتم اختيارها في أغلب الأحيان. علاوة على ذلك، من المحتمل أن يختار من هم في النهاية العليا (أقصى اليمين) الخيار 3 (“معظم الوقت”)، في حين أن الموجودين في الطرف السفلي (أقصى اليسار) سيختارون الخيار 1 (“نادراً”). وبالتالي، حتى ضمن هذا النطاق الضيق، لا يزال بإمكاننا إجراء تمايز بين الأشخاص الذين هم أعلى وأقل فيما يتعلق بالدوخة.



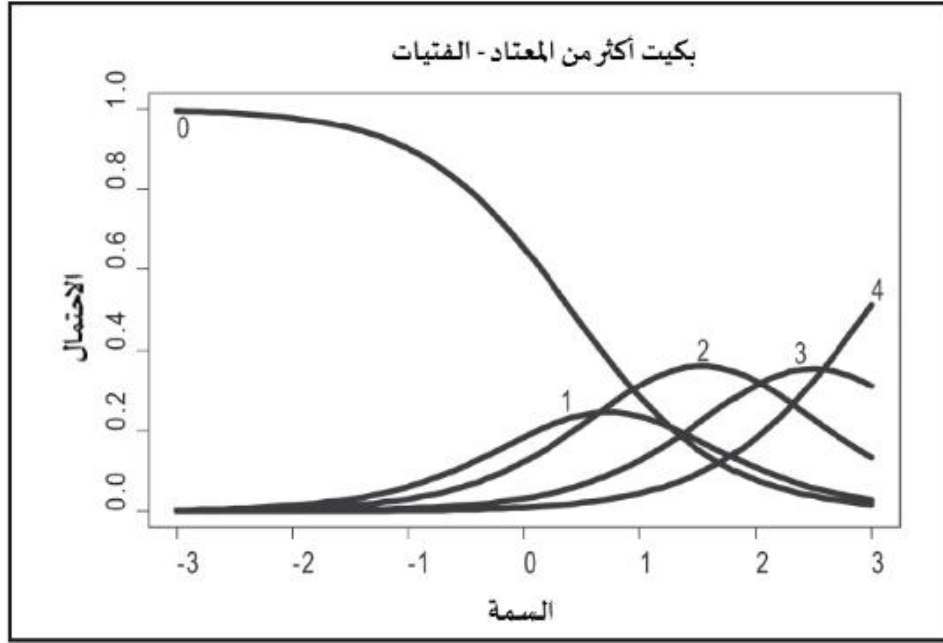
شكل (٧-٨): مثال افتراضي لمنحنيات استجابة الفئة مزدحمة عند الطرف السفلي (الجهة اليسرى) من مقياس السمة

في المقابل، إذا نظرنا إلى موقع أبعد على مقياس السمة، في حدود ما بين 0.2 و 0.3، فإن الاحتمال الغالب هو أن المستجيبين الذين لديهم مستويات دوار في أي موقع بين تلك القيمتين سيختارون الخيار 4 («تقريباً دائماً»). في الواقع، «تقريباً دائماً» هي الاستجابة الأكثر احتمالاً لأي مستجيب يحقق أعلى من 8.0 على مقياس السمة. وبالتالي، يوفر هذا البند تمييزاً بسيطاً بين الأشخاص الذين يحققون ما بين 8.0 و 0.3 بالنسبة لمستويات الدوخة لديهم. منحنيات الاستجابة للبند في الشكل (7-8) لديها الكثير لتخبر عنه. لاحظ أن منحنى استجابة الخيار 2 («بعض الوقت») تتداخل قليلاً مع منحنى الاستجابة للخيار 1 («نادراً») والخيار 3 («معظم الوقت»). لا يوجد فعلياً جزء من نطاق القدرة الذي يكون فيه خيار الاستجابة 2 «بعض الوقت» هو الأكثر احتمالاً للاختيار. وبشكل جوهري، فإن خيار الاستجابة 2 «بعض الوقت» لا يقوم بأي شيء لا يمكن لخاري الاستجابة الواقعيين على جانبيه أن ينجزاه. لا يزال من المرجح أن يقوم المستجيب الذي حصل على درجة من الدوخة الصباحية تعادل الدرجة (-1, 0) على مقياس السمة (النقطة التي من المرجح أن يختار المجيبون عليها خيار الاستجابة 2 باختيار الخيار 1 أو 3. إن اختيار الخيار

2 في هذه الحالة الافتراضية يؤدي إلى الغموض بين الدوخة النادرة والدوخة في بعض الأحيان. بالأخذ في الاعتبار هذا البند بشكل منعزل، وبغض النظر عن أي بنود أخرى قد يتم تضمينها مع هذا البند في المقياس، فإن إسقاط خيار الاستجابة 2 «بعض الوقت» من المحتمل أن يكون له معنى.

شكل المنحنيات لخيارات الاستجابة يعطينا معلومات حول خاصية التمييز discrimination. وهذا يبدو أكثر وضوحاً بالنسبة للمنحنى الأخير، والذي يشبه إلى حد كبير المنحنيات التي قمنا بفحصها المتعلقة بنموذج الاستجابة الثنائية (اجتياز/فشل) ذي الثلاث معلمات (كما هو موضح في الشكل (7-3)). المنحنى الموجود في أقصى اليسار هو نوع من صورة معكوسة طبق الأصل لما رأيناه في نموذج المعلمات الثلاث من حيث إنه يصل إلى أعلى مستوياته في أقصى اليسار وينحدر مع التقدم للأعلى في مقياس السمة، ولكن مرة أخرى، فإن انحدار ميل المنحنى يدل على درجة تمييزه للسمة. بالنسبة للمنحنيات الموجودة بين المنحنيات الطرفية، يشير الانحدار (أو ربما بدقة أكثر، التفرطح kurtosis) إلى درجة التمييز. خيارات الاستجابة التي تكون منحنياتها مرتفعة ومدببة تميز بشكل أفضل من تلك التي تكون منحنياتها أكثر اتساعاً وانبساطاً. في الشكل رقم (7-8)، خيار الاستجابة 2 («بعض الوقت») له منحنى على شكل القرفصاء إلى حد ما، مما يشير إلى أن خيار الاستجابة لا يميز بشكل جيد. هذا يدعم كذلك فكرة أن حذف هذا الخيار قد يكون مناسباً.

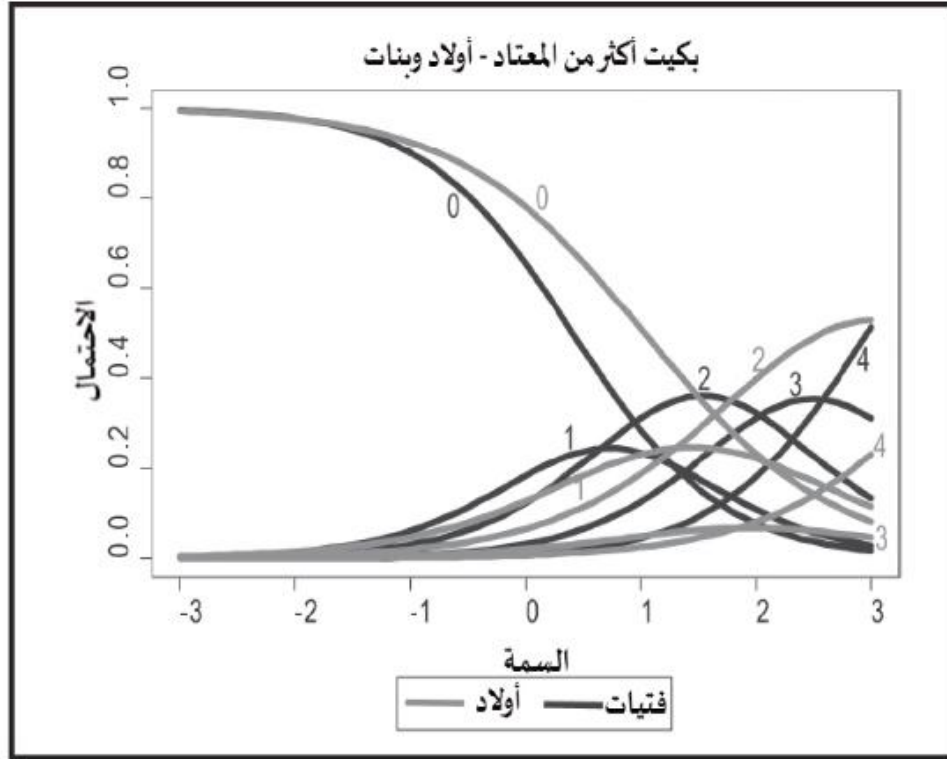
لاختتام هذا القسم، سننظر في منحنيات الاستجابة للفئات (الخيارات) الناتجة عن بيانات فعلية (تم توفيرها واستخدامها بإذن من زملائي دارين دي ولت Darren DeWalt وديفيد ثيسن David Thissen من جامعة UNC) وسوف نستخدم هذه المنحنيات لتوضيح الغرض الآخر الذي يمكنها تقديمه. يمثل الشكل رقم (7-9) بنداً تم تطبيقه على مجموعة من المرضى الأطفال من الإناث كجزء من جهد لتطوير مجموعة بنود لتقييم الاكتئاب. هذا البند يطلب من الأطفال الإشارة إلى عدد المرات التي بكوا فيها أكثر من المعتاد في الماضي القريب. على نحو مشابه لكن غير مطابق لأمثلتنا السابقة، كانت خيارات الاستجابة هي صفر («أبداً») و 1 («أبداً تقريباً») و 2 («أحياناً») و 3 («غالباً») و 4 («دائماً تقريباً»). يشير تجمع المنحنيات إلى اليمين إلى أن المشاركين يجب أن يكونوا على درجة عالية من السمة (أي الاكتئاب) من أجل تقديم إجابة بخلاف «أبداً».



شكل (٧-٩): منحنيات استجابة الفئة للفتيات للبند المتعلق بالبكاء أكثر من المعتاد

يبين الشكل رقم (٧-١٠) استجابات الأولاد واستجابات الفتيات معاً على نفس الشكل. على الرغم من أن هذا الشكل مزدحم للغاية، إلا أنه يكشف أن منحنيات الأولاد، والفتيات متشابهة عمومًا ولكنها ليست متطابقة بأي حال من الأحوال. ولجعل الفارق الحاسم بين استجابات الأولاد والفتيات أكثر وضوحًا، في الشكل رقم (٧-١١)، تم تحرير المنحنيات التي تمثل خيارات الاستجابات الوسطى الثلاثة، وتم الإبقاء فقط على تلك المنحنيات الخاصة بخيارات الاستجابة صفر "أبدأ" و 4 "دائمًا تقريبًا". بالنسبة للخيار صفر، يكون منحني الأولاد أعلى من الفتيات، أما بالنسبة للخيار 4، فإن الفتيات "أعلى من الأولاد".

كيف يجب أن نفسر هذا الاختلاف؟ هذا الاختلاف يشير إلى أنه خلال المدى الكامل لقيم السمة (الاكتئاب)، فإن الأولاد أكثر أرجحية من الفتيات للقول إنهم لم يبكوا (أي اختيار خيار الاستجابة "أبدأ") أكثر من المعتاد. في الطرف العلوي من مقياس السمة، تكون الفتيات أكثر احتمالية من الأولاد للاعتراف بأنهم يبكوا أكثر من المعتاد "دائمًا تقريبًا" خلال الفترة الزمنية المعنية.



شكل (٧-١٠): منحنيات استجابة للفئة متداخلة للفتيان والفتيات للبند المتعلق بالبكاء أكثر من المعتاد

يشير هذا النمط إلى أحد أمرين: إما (1) أن تكون الفتيات أكثر اكتئاباً من الأولاد أو (2) يوّلد مستوى منخفض من الاكتئاب استجابات أعلى لدى الفتيات منها لدى الأولاد. أي إن البند يعكس الفروق الحقيقية بين الجنسين بين الأولاد والفتيات أو أن البند لا يعمل بشكل متكافئ بين الجنسين. الشك لدينا هو أن الإقرار بالبكاء قد يكون أكثر صعوبة لدى الفتيان منه لدى الفتيات. ونتيجة لذلك، فإن الأولاد الذين يختارون أي خيار استجابة بخلاف «أبداً» قد يكونون بالفعل أكثر اكتئاباً من الفتيات اللاتي يخترن خيار الاستجابة نفسه. وهذا يشير إلى أن البند يظهر أداء البند التفاضلي (DIF) differential functioning item. بمعنى أن المجموعات المختلفة من المجيبين (الأولاد، والفتيات في هذه الحالة) الذين لديهم مستويات مماثلة من السمة يحصلون على درجات مختلفة للبند، مما يشير إلى أن البند يحدد على مقياس السمة بشكل مختلف كدالة في مجموعة الجنس. من خلال النظر إلى البنود الأخرى التي تقيس الاكتئاب والتي لا تسأل عن البكاء ومن خلال ملاحظة النتائج المتشابهة بشكل أساسي للبنين والبنات، وجدنا دليلاً إضافياً على أداء البند التفاضلي DIF للبند المعني. بينما توجد طرق أخرى لتحديد أداء البند التفاضلي DIF، يجد



العديد من الباحثين المعلومات المرئية التي تنتجها حزم نظرية الاستجابة للبند IRT لتكون مقنعة بشكل خاص.

خلاصة القول، يمكن أن يوفر فحص منحنيات استجابة الفئات معلومات من عدة أنواع حول كيفية أداء بند ما. ومن بين الأشياء الكثيرة التي يمكن أن يكشف عنها فحص منحنيات الاستجابة للفئة (أ) معلومات حول مدى التمثيل الكلي لبند ما للمجموعة الكاملة للخاصية التي يتم قياسها، (ب) ما إذا كانت خيارات الاستجابة المرتبة تمثل بدقة المناطق الأعلى من مقياس السمة بشكل تدريجي، (ج) ما إذا كانت جميع خيارات الاستجابة ضرورية ومفيدة، (د) مدى قدرة البند على التمييز بشكل جيد عند مستويات مختلفة من السمة، و(هـ) ما إذا كان البند يظهر أداء البند التفاضلي DIF. يمكن الحصول على أنواع أخرى من المخرجات الرسومية من العديد من حزم نظرية الاستجابة للبند، ويكون التمثيل المرئي لأداء العنصر من بين أكثر ميزات IRT فائدة.

### ثيتا والاختبار التكيفي المحوسب (Theta and Computerized Adaptive Testing) (CAT):

في قلب كل أساليب القياس، هناك مفهوم أن قوة السمة التي يتم قياسها تختلف باختلاف الأفراد أو الوقت، ونظرية الاستجابة للبند IRT ليست استثناءً من ذلك. وفي وقت سابق من هذا الفصل، تم استخدام مصطلح "القدرة" لوصف مثل هذه السمة. وغالبًا ما يشار إلى هذه السمة، بصرف النظر عن محتواها الموضوعي، بـ "ثيتا" في سياق نظرية الاستجابة للبند IRT ويمكن الإشارة إليها بالحرف اليوناني الكبير  $\theta$ . كما ذكر سابقًا في هذا الفصل، يمكن إنشاء البنود بحيث يكون لها مستويات مختلفة من الصعوبة. وبشكل أساسي، تتم معايرة الصعوبة من حيث ثيتا. وبالتالي، فإن الاستجابة الإيجابية لبند أكثر صعوبة (أو اختيار خيار استجابة يمثل قيمة أعلى) تعني أن المستجيب لديه قيمة ثيتا أعلى (أي أنه يمتلك مستوى أعلى من السمة التي يقيّمها البند) نسبيًا للمستجيبين الذين يؤيدون فقط البنود المرتبطة بقيم ثيتا الأقل. بالنسبة لكل بند، يمكن لمحلل البيانات تحديد مستوى (أو نطاق) قيم ثيتا التي يميز البند عندها بين المشاركين بمستويات مختلفة من السمة. بعض البنود قد تعمل بشكل جيد عبر مجموعة واسعة من قيم ثيتا، بينما بنود أخرى ستعمل بشكل جيد فقط عبر نطاق ضيق من ثيتا.

إن بنداً افتراضياً مثل «أنا قلق على مستقبلي» مقروناً بخيارات الاستجابة (0) «أبداً» (1) «القليل من الوقت» (2) «معظم الوقت» و(3) «دائماً»، قد يكون قادراً على توفير لمحة عامة للسمة التي يقيّمها عبر نطاق واسع من ثيتا. ومع ذلك، قد لا يفرق بشكل جيد للغاية بين التقديرات الأصغر للخاصية، مثل نادراً rarely ما تشعر بالمحتوى وقلماً seldom (بمعنى أكثر بقليل من نادراً) تشعر بالمحتوى. قد يختار الأشخاص في كل من هذين المستويين الأخيرين من القلق على المستقبل «القليل من الوقت»، وبالتالي لا يمكن للبند تمييز الفرق الدقيق في «القلق على المستقبل» بينهما. ولكن بالنسبة لبعض التطبيقات، قد يكون التمييز بين الأفراد أصحاب الخيارات الدقيقة «نادراً ما» rarely و«قلماً» seldom يعتبر أمراً مهماً. وبالطبع، يمكن إنشاء بنود إضافية تركز بمزيد من التفصيل على جزء أقصر من النطاق الكامل للسمة. قد يكون لهذا البند خيارات إجابة مثل «أبداً»، و«نادراً جداً»، و«نادراً»، و«قلماً»، و«من حين لآخر». أو قد تتم إعادة صياغة أصل البند نفسه لاستهداف مستوى مختلف من ثيتا. على سبيل المثال، «لدي شعور بالفزع حيال مستقبلي» يثير بوضوح المخاطر. أي من هذين النموذجين البديلين للبند قد يركّز على المستويات الأعلى لسمة القلق أكثر من البند الأصلي.

في الواقع، يمكن تطوير مجموعة من البنود من هذا النوع تستهدف أجزاء مختلفة من سلسلة قيم ثيتا، بعضها يركز في المقام الأول على المستويات المنخفضة، والبعض الآخر على المستويات المتوسطة، والبعض الآخر على مستويات عالية من القلق. من المحتمل أن تتفوق مجموعة كبيرة ومصاغة بشكل متقن من هذه البنود على عدد أقل من البنود العامة. وهذا يعني أن هذه المجموعة الموسعة من البنود يمكن أن تحل الاختلافات الأصغر في ثيتا بدقة أكبر (أي تفريق المجيبين الذين كانت مستويات قلقهم أكثر تشابهاً) بالإضافة إلى تمديد النطاق الكامل لثيتا (أي، تضمين بنود مستهدفة لعدة مستويات مختلفة من القلق تتراوح من منخفض للغاية إلى مرتفع للغاية). المشكلة في هذه الإستراتيجية هي أنها تتطلب عدداً كبيراً من البنود. علاوة على ذلك، قد يضطر أي مجيب معين إلى إكمال عدد كبير من البنود التي كانت خارج نطاق ثيتا الخاص به (على سبيل المثال، يتعلق ذلك بمستويات أعلى أو أقل من السمة من تلك التي يمتلكها المجيب). هذا من المحتمل أن يكون له ضريبة على المجيبين الذين قد يشعرون بالضجر من استكمال العديد من البنود والذين ربما يكونون منزعجين من البنود التي لا تنطبق على ظروفهم. إن إدراك هذه المشاكل المحتملة قد يلزم مطوري المقياس إما باختيار مجموعة أقصر من البنود التي قد لا تؤدي مهمة جيدة لاكتشاف

الاختلافات الأكثر دقة في ثيتا، أو باختيار أداة أطول تميز بدقة أكبر ولكن من المحتمل أن تكون أكثر كلفة.

يقدم الاختبار التكيفي المحوسب (CAT) علاجًا ممكنًا لهذه المعضلة. في CAT، تم تطوير بنك كبير من البنود، وتتم معايرة كل منها بعناية بحيث تكون قيمة  $\theta$  التي تتوافق معها معروفة. يكمل المحببون البنود عبر الكمبيوتر، مع تحديد خوارزمية مسبقة الضبط بشكل تكيفي لتحديد عدد البنود المعروضة وترتيبها. البنود القليلة الأولى عادةً ما تكون عامة إلى حد ما، ولا تسفر إلا عن فكرة تقريبية عن مستوى المحبب من السمة التي يجري تقييمها. استنادًا إلى الإجابات على تلك البنود الأولية، يمكن لخوارزمية الكمبيوتر تقديم البنود التالية بشكل انتقائي، واختيار تلك البنود الموجودة ضمن مجموعة البنود المتوفرة الأكثر صلة والغنية بالمعلومات لمستوى ثيتا التي يبيدها المستجيب المحدد. بمعنى، يمكن للكمبيوتر تقديم البنود التي تغطي منطقة من السلسلة المتصلة لقيم ثيتا القريبة من مستوى السمة التي كشفت عنها الاستجابات السابقة. لذلك، على سبيل المثال، إذا كانت الاستجابات الأولية تشير إلى أن المحبب يقلق بشأن مستقبله أكثر من معظم الأشخاص، فيمكن اختيار البنود التي تركز على مستويات أعلى من القلق بينما يمكن حذف البنود التي تستهدف مستويات منخفضة للغاية. مع استمرار التقييم، يمكن للكمبيوتر تحسين اختيار البنود، مع التأكد من أن البنود المعروضة ذات صلة بمستوى سمة المحبب وكافية في العدد لتوفير وصف دقيق.

بالمعنى الحرفي للكلمة، لا يعتمد الاختبار التكيفي من هذا النوع على استخدام نهج نظرية الاستجابة للبند IRT. في الواقع، قبل تطوير هذه الأساليب بفترة طويلة، استخدمت اختبارات الذكاء الموحدة استراتيجية مماثلة لطرح بعض الأسئلة "محددة الموقع" («locator» questions) لتحديد أي من الأسئلة المتعلقة بمنطقة معينة من الذكاء يجب طرحها وتحديد عددها. ومع ذلك، فإن القدرة التي يوفرها نهج IRT لمعايرة البنود لمستويات محددة من ثيتا (وللتعبير عن هذه المعلومات من الناحية الكمية)، إلى جانب قدرة أجهزة الكمبيوتر على الاستفادة في الوقت الفعلي من مجموعة الاستجابات المقدمة واستخدام تلك البيانات لتحديد البند التالي الذي سيتم تقديمه، زادت بشكل كبير من فائدة واعتماد الاختبار التكيفي.

الاختبار التكيفي المحوسب CAT يمكن أن يكون جذاباً للغاية عندما يكون ذلك ممكناً. وهو يتطلب وجود بنك كبير من البنود التي تم اختبارها بشكل كافٍ لتوفير معايرة جديرة بالثقة. أيضاً،

تحتاج خوارزميات CAT إلى حساب خصائص البند بخلاف نتائج المعايرة. على سبيل المثال، قد يؤدي الاختيار على أساس ثبوتنا إلى التوضيح دون قصد بنطاق المحتوى إذا تم استبعاد البنود التي تمثل بنوداً مهمة للمفهوم البنائي (construct) قيد الفحص. لا تزال الدراسات التجريبية التي تقارن CAT مع استخدام مجموعات البنود الثابتة نادرة نسبياً، على الرغم من أن التقارير المتاحة تشير إلى وجود ميزة لـ CAT مقارنةً بمجموعات البنود الثابتة ذات الحجم المتواضع (مثل Cella & Hays, Pilkonis, Reise, Choi, 2010). يبدو من المحتمل أنه كلما كان نطاق المحتوى موضع اهتمام أكبر، وكلما كانت مستويات السمة أكثر تقلباً بين شرائح مختلفة من المجتمع، فإنه من الممكن أن تكون فوائد الاختبار التكيفي المحوسب CAT أكبر.

للحصول على نظرة أكثر شمولاً على الاختبار التكيفي المحوسب CAT، يمكن للقراء الرجوع إلى (Bjorner, Chang, Thissen, and Reeve (2007). المجلد الأخير الذي يمكن للقراء استخدامه لتعزيز فهمهم لمجموعة واسعة من موضوعات نهج نظرية الاستجابة للبند IRT هو كتيب نمذجة نظرية الاستجابة للبند: تطبيقات لتقييم الأداء النموذجي Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment (Reise & Revicki, 2015).

### تعقيدات نظرية الاستجابة للبند IRT:

على الرغم من أنها جذابة للغاية، فإن نظرية الاستجابة للبند IRT ليست حلاً سريعاً لمشاكل القياس. مثل نظرية القياس الكلاسيكية CTT، لا تحدد IRT خصائص البنود؛ إنها مجرد قياس كمي للبنود. لذا، فإن التكنولوجيا في حد ذاتها تسمح للباحث بتقييم أداء البند، لكنها لا تتسبب بشكل مباشر في كتابة بنود أفضل أو جعل البنود التي تم إنشاؤها بشكل سيء تعمل فجأة بشكل جيد. أيضاً، يمكن أن تكون عملية التقييم شاقة عند استخدام طرق تعتمد على IRT. القياس الكلاسيكي يضحى بالدقة لأجل البساطة من خلال تبني تصورات أقل تبايناً ولكن أكثر تعقيداً لمصادر الخطأ. بينما تقوم نظرية الاستجابة للبند IRT بالخيار العكسي، واكتساب الدقة مقابل التوضيح بالبساطة. علاوة على ذلك، يتطلب تطبيق طرق IRT درجة كبيرة من قرار الخبراء. لا تزال هذه الطرق في مرحلة التطوير النشطة، مع ظهور مشكلات جديدة وتقديم حلول جديدة.

واللحصول على الثقة بأن خصائص البنود يتم تقييمها بشكل مستقل عن خصائص العينة التي تتم دراستها، وهو الهدف الأساسي لـ IRT، يجب على المرء أن يثبت أن هذه الخصائص متسقة عبر مجموعة واسعة من المشاركين الذين يختلفون في نواح مختلفة، بما في ذلك مستوى القدرة. ومن المهم ألا ترتبط خصائص البند بخصائص العينة الأخرى المستقلة عن السمة، مثل الجنس أو العمر أو المتغيرات الأخرى التي يجب أن تكون غير مرتبطة مع تلك التي يتم قياسها. يجب أن تختلف درجات البنود فقط عندما تتغير السمة محل الاهتمام، وليس بسبب الاختلافات في أي متغيرات خارجية المنشأ extraneous variables. لذلك، على سبيل المثال، إذا افترضنا أن القدرة الإملائية spelling لا علاقة لها بالجنس، فسيتم علينا إثبات أن الأولاد والبنات ذوي القدرة المتساوية سيكون لديهم نفس الاحتمال في اجتياز بند ما. إذا لم يكن هذا صحيحًا، فسيؤثر نوع الجنس أو بعض العوامل الأخرى بخلاف القدرة الإملائية على البند. أيضًا، كما هو الحال مع النظرية الكلاسيكية، يجب أن تشترك البنود التي يتم فحصها في مجموعة (أي تكوين مقياس لقياس المتغير) في متغير أساسي واحد فقط.

إحدى نتائج هذه المتطلبات هي أنه يجب على الباحثين الوصول إلى أعداد كبيرة من المجيبين المتنوعين من أجل وصف البنود بدقة. الهدف الرئيسي لـ IRT هو ربط البنود بمستويات محددة من السمة التي يتم تقييمها. البنود الجيدة والعينات الكبيرة ضرورية لتحقيق ذلك. تتطلب الاستقلالية النظرية لخصائص البنود وخصائص العينة التي تعد سمة مميزة لـ IRT أن يكون قد تم تقييم البنود عبر النطاق الكامل للسمة محل الاهتمام وأن أداة القياس تتضمن بنوداً حساسة بشكل شامل لجميع مستويات تلك السمة.

كيف يتحقق مطور أدوات القياس من المستوى الحقيقي للسمة بطريقة تسمح بتكوين المنحنيات المميزة للبنود ICCs؟ بالعودة إلى تشبيه المطرقة والجرس، كيف يمكنك تحديد القوة من أجل تحديد مقدار القوة التي يتطلبها قرع الجرس لطرق جزئية معينة من الجهاز؟ في كثير من الحالات، إذا كان المستوى الحقيقي للسمة معروفًا بطريقة يمكن التحكم فيها، فلن تكون هناك حاجة كبيرة لتطوير مقياس جديد. من الناحية النظرية، نظرًا لاستجابات عدد كبير من الأشخاص لمجموعة ثابتة من البنود، يجب أن يكون برنامج الكمبيوتر قادرًا على فرز الاختلافات الناشئة عن خصائص البند مقابل خصائص الشخص. للعودة مرة أخرى إلى تشبيه أجهزة الكرنفال (جهاز

رنين الجرس وآلة الغمر)، إذا استخدم عدد كافٍ من الأشخاص كلاً من الجهازين، على سبيل المثال، يجب أن يكون من الممكن تحديد أي من الجهازين أصعب، وأيضاً، للحكم على مهارات الأفراد في هاتين المهمتين. في الممارسة العملية، غالباً ما تكون هناك عملية تكرارية متداخلة تتضمن إدارة البنود لقياس مستوى السمة لمجيبين معينين، ثم استخدام تقدير السمة كدليل في تحديد خصائص البنود الأخرى. عندما يتم تحديد أفضل البنود على هذا الأساس، يمكن استخدامها للحصول على تقدير محسّن لمستويات الأفراد في السمة للجولة التالية من اختيار البنود، وهكذا.

نظراً لطبيعة هذه العمليات، من السهل أن نرى السبب وراء تبني نظرية الاستجابة للبنود IRT بحماس شديد من قبل المؤسسات التجارية المعنية باختبار القدرات، مثل تلك التي تدير اختبارات سجل الخريجين. يوفر التطبيق والتقييم المستمر للبنود بمرور الوقت أساساً ممتازاً للعثور على البنود التي تنسم خصائصها بالاستقرار خلال التباين في مجموعة واسعة من خصائص المستجيبين الأخرى.

أحد التعقيدات الأخرى لنظرية الاستجابة للبند IRT يتعلق باستخدام بنود مختلفة لقياس الشيء نفسه في أوقات مختلفة. في وقت سابق، ذكرت باختصار الاختبار التكيفي المحوسب (Van der Linden & Glas, 2000) (CAT) كنهج لجمع البيانات المرتبطة غالباً بـ IRT. في CAT، تتم معايرة البنود بشكل فردي لمقياس السمة باستخدام طرق IRT، وبالتالي، يمكن اختيارها للتطبيق بحيث تتوافق مع مستوى السمة لمجيب معين. هذا يعني أن البنود المختلفة ستكون مناسبة على النحو الأمثل لجوانب القوة المختلفة للسمة. يمكن أن تحدث مثل هذه الاختلافات إما بين الأشخاص الذين يختلفون فيما يتعلق بالسمة أو لفرد معين يتغير مستوى السمة الخاصة به.

وبالتالي، فإن استخدام الاختبار التكيفي المحوسب CAT غالباً ما يستلزم نتائج أو درجات من بنود مختلفة تتم مقارنتها مع بعضها البعض. على سبيل المثال، إذا تغيرت إحدى السمات كنتيجة للتدخل، فقد تختلف أفضل البنود لقياس تلك السمة في أي فرد بعد التدخل عن ذي قبل. بالنسبة للمستخدم لتقرير علمي يبني ادعاء نجاح التدخل على أساس مقارنة بين البنود المختلفة عبر (بسبب الاختلافات الأولية في السمة) وداخل (بسبب التغييرات الناتجة عن التدخل) الأفراد، فإن الوثوق في البيانات قد يتطلب الإيمان بذلك. يتمتع معظمنا بخبرة أكبر بكثير (وربما بالراحة) في الظروف التي تم فيها إعطاء البنود نفسها قبل وبعد التدخل مقارنةً بالسيناريو الذي تم وصفه للتو.

عندما يتم تطبيق طرق IRT بشكل صحيح وتتم معايرة العناصر بعناية بالنسبة لسلسلة قيم السمة على أساس البيانات التي تم جمعها من عينات كبيرة ومتنوعة، فإنه يمكن تحويل الدرجات الناتجة من بنود مختلفة إلى مقياس سمة شائع بنفس الطريقة التي يمكن من خلالها تحويل قياسات بالبوصة أو بالمليمتر إلى مقياس شائع للطول. عندما تكون الدقة التي يمكن بها تعيين البنود على بُعد السمة منخفضة، فإن تفسير النتائج بناءً على مجموعات بنود مختلفة قد يكون مشكلة. علاوة على ذلك، نظرًا لأن مستخدمي تقارير الأبحاث قد لا يفهمون تمامًا جميع المشكلات التي ينطوي عليها نوع مقارنات البنود المختلفة التي قد ينطوي عليها الاختبار التكيفي المحوسب CAT، فقد يواجه مؤلفو هذه التقارير درجة أكبر من الشك.

#### الخلاصة:

إن أساليب القياس القائمة على نظرية الاستجابة للبند IRT لديها العديد من الميزات الجذابة. ولكن إنشاء البنود الجيدة هو عمل شاق، بغض النظر عن الإطار النظري الذي يوجه العملية. إن كتابة البنود التي تقيس السمة قيد الدراسة باتساق والتي لا تراعي خصائص المستجيب الأخرى ليست مهمة سهلة. بينما في المقاييس التي تعتمد على النظرية الكلاسيكية، فإن وجود بنود متعددة يمكن أن يعوض عيوب بعض هذه البنود إلى حد ما، في حين أن المنطق وراء نظرية الاستجابة للبند أن كل بند يتم الحكم عليه على حده. بسبب أنه يمكن للمرء أن يكتشف، من خلال فحص المنحنيات المميزة للبنود (ICCs)، على سبيل المثال، كون أداء أحد البنود جيدًا لا يعني أن ذلك المرء سيفعل ذلك. إن امتلاك معرفة مستقلة موثوقة بالخصائص التي يتم قياسها هو أحد متطلبات نظرية الاستجابة للبند التي يصعب الوفاء بها بدقة، ولكن يمكن تقريبها بشكل كافٍ بالاختبارات المتكررة للعينات الكبيرة وغير المتجانسة. وبالتالي عندما لا يكون هذا خيارًا، فقد يكون من الصعب للغاية إقناع النقاد بأن الافتراض قد تحقق بشكل كافٍ.

وجهة نظري الشخصية هي أنه عندما تكون افتراضات نظرية القياس الكلاسيكية CTT قابلة للتطبيق (أي، حيث يتم تصميم البنود كمؤشرات مكافئة لمتغير أساسي مشترك)، فإن قابلية التنبع والأداء لهذه المقاييس تجعلها خيارات جذابة. من ناحية أخرى، في حالة توفر عينات كبيرة (والتي يُعد عنها الاختبار التكيفي المحوسب CAT طريقة مفضلة للتطبيق)، أو في حالة أن الدرجات المكافئة في دراسات متعددة تكون ذات أهمية قصوى، فإن الأساليب القائمة على IRT في

ظل وجود التعقيد المضاف قد تكون هي الخيار الأفضل. ومع ذلك، فإن مجرد استخدام هذه الأساليب ليس ضماناً للمنتج النهائي المنشود. ويجب على الباحث أن يثبت أن افتراضات الطريقة المختارة قد تم الوفاء بها ضمن حدود مقبولة وأن ثبات reliability أداة القياس الناتجة وصدقها validity قابلة للتحقق منها تجريبياً.

هل تجعل نظرية الاستجابة للبند الأساليب الكلاسيكية قديمة؟ يدرك العديد من مؤيدي نظرية الاستجابة للبند أن لكل من نظرية القياس الكلاسيكية ونظرية الاستجابة للبند دور تلعبه. على سبيل المثال، ذكر (Embretson & Hershberger (1999، في أول توصياتهما المتعلقة بالتغيرات في أساليب القياس الحالية، أنه يجب دمج نظرية الاستجابة للبند والأساليب الكلاسيكية CTT في نهج شامل لقضايا القياس. (ص 252).

على الأقل، خلصت دراستان تجريبيتان كبيرتان (Fan, 1998; Stage, 2003) التي قارنت خصائص الاختبارات التربوية القائمة على الأساليب الكلاسيكية ونظرية الاستجابة للبند إلى أن الأساليب الكلاسيكية أدت بشكل جيد أو أفضل من الأساليب القائمة على نظرية الاستجابة للبند. استندت دراسة (Fan (1998 إلى 40 عينة كل منها مكونة من 1000 فرد، مأخوذة من مجموعة تضم حوالي 193000 تلميذ من تكساس تم اختبارهم في مهارات القراءة والرياضيات. وقد أظهرت خصائص البند والشخص اتساقاً كبيراً عبر الأساليب. قامت (Stage (2003 بمقارنة الطرق التقليدية وطرق نظرية الاستجابة للبند باستخدام بيانات من اختبارات القدرات القياسية السويدية Swedish Standard Aptitude Tests. واستناداً إلى عينة من 2441 من المتقدمين للاختبارات تم اختيارهم عشوائياً من مجموعة من 82506، خلصت إلى أنه في حين أن نموذج نظرية الاستجابة للبند المكون من ثلاثة معلمات كان متوافقاً (fit) مع البيانات بشكل ضعيف، فإن نموذجاً قائماً على نظرية الاختبار الكلاسيكية كان جيداً. في الآونة الأخيرة، أجرى Silvestro (2009) دراسة مماثلة ولكنها أصغر باستخدام درجات اختبار لـ 326 طالباً جامعياً من طلبة السنة الأولى. وقد خلص إلى أن:

النتائج هنا توضح ببساطة أن إطار القياس ينتجان إحصائيات متماثلة لكل بند وشخص سواء من حيث قابلية المقارنة بين إحصاءات البند والشخص، ومستوى صعوبة البنود، والاتساق الداخلي، وأداء البند التمييزي بين إطار العمل. (ص 29).



وعلى نفس المنوال، قدم باحثان من خدمة الاختبارات التربوية ومجلس الكلية مؤخرًا ورقة تصف نتائج دراسة محاكاة قارنت طرق نظرية الاستجابة للبند والأساليب الكلاسيكية مع تحديد نماذج الاختبار (Davey & Hendrickson, 2010). وخلصوا إلى أن "كلا النهجين يوفران مزيجًا مثيّرًا للاهتمام من المزايا والعيوب النظرية والعملية، مع عدم ظهور أيٍّ منهما كمفضل واضح على الورق". وقد لاحظوا كذلك أنه:

كانت النتيجة الأكثر وضوحاً للدراسة هو الفرق البسيط نسبياً بين أداء طرق الكلاسيكية ونظرية الاستجابة للبند .... قد يعني التشابه في الأداء لتلك الأساليب أن اختيار الممارس يعتمد إلى حد كبير على التفضيل أو الملاءمة التشغيلية. (ص 2)

لا ينبغي تفسير هذه التقارير على أنها تشير إلى أن أساليب نظرية الاستجابة للبند ليست لها مزايا في أي موقف. لا شيء من هذه الدراسات قطعي، وبعض هذه الاختلافات تم تحديدها عند مقارنة تلك الطرق. تحليلات نظرية الاستجابة للبند في بعض الحالات قدمت معلومات أكثر تفصيلاً من التحليلات الكلاسيكية. وكما رأينا في الرسوم البيانية للبنود متعددة الاستجابات، يمكن لطرق نظرية الاستجابة للبند أن تقدم رؤى مفيدة، مثل الإشارة إلى أن جميع خيارات الاستجابة ليست ضرورية أو أن البند لا يلتقط معلومات عبر النطاق الكامل للسمة التي يتم قياسها. توفر طرق نظرية الاستجابة للبند أيضاً ميزات نظرية واضحة (مثل استقلال خصائص البند عن خصائص العينة)، ومع تحسن الأساليب، قد يتم تقليص بعض العوائق العملية. ولكن كحد أدنى، تجادل هذه الدراسات أن طرق نظرية الاستجابة للبند ليست بالضرورة أفضل من طرق القياس الكلاسيكية. باختصار، إن تقارير زوال القياس الكلاسيكي سابقة لأوانها.

ولتناول مزايا وعيوب كلا الطريقتين، لاحظ (Zickar and Broadfoot (2008):

كما أن طرق نظرية القياس الكلاسيكية CTT لها قيودها الخاصة، لاحظ الباحثون وجود قيود شديدة على طرق نظرية الاستجابة للبند IRT التي تجعل استخدام أساليبها صعبة أو مستحيلة أو غير عملية في سيناريوهات معينة. تتضمن هذه القيود الحاجة إلى أحجام كبيرة للعينات، وافتراضات قوية لأحادية البعد unidimensionality [وهو ما ينطبق أيضاً على الأساليب الكلاسيكية]، وصعوبة تنفيذ البرامج. (ص 48).

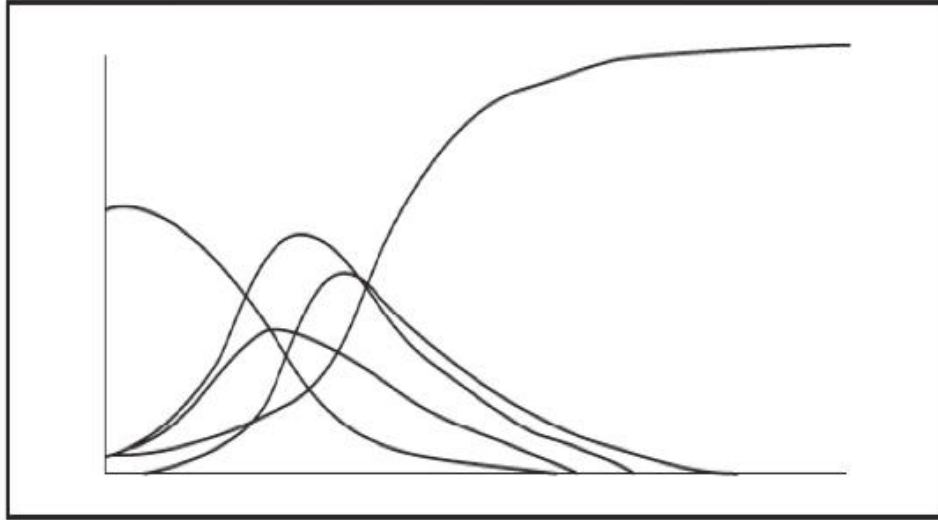
فيما بعد في نفس التقرير، يذكر هذان المؤلفان:

على الرغم من أن الأسطورة الشائعة أن نظرية القياس الكلاسيكية قد ماتت، إلا أننا نعتقد أن هناك العديد من السيناريوهات التي فيها يفضل استخدام نظرية القياس الكلاسيكية. معظم الأسباب التي يمكن تصنيفها تعود للقيود في البيانات التي قد تستبعد نظرية الاستجابة للبند والاعتبارات العملية التي قد تجعل نظرية القياس الكلاسيكية أكثر تفضيلاً. (ص 50).

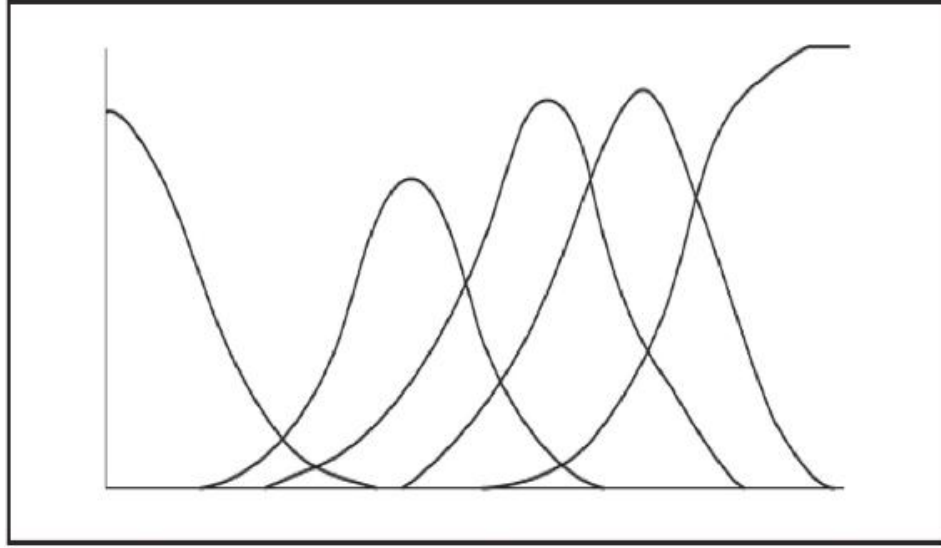
الطرق القائمة على نظرية الاختبار للبند IRT ستستمر شعبيتها في الزيادة. وسيكون لها مزايا واضحة على الطرق السابقة في ظروف معينة. سوف تتعايش نظرية الاستجابة للبند والأساليب الكلاسيكية، بقدر مشاركة تحليل الانحدار للمرحلة مع أساليب نمذجة المعادلات البنائية. على الرغم من أن كلاً من نظرية الاستجابة للبند ونمذجة المعادلات البنائية حملتا أشياء أبعد من سوابقهما، إلا أن الطرق السابقة احتفظت بفائدتها.

تمارين:

1- تمثل الأشكال التالية نتائج تحليل نظرية الاستجابة للبند لاثنتين من البنود الافتراضية، ولكل منها خمسة خيارات استجابة.



شكل (٧-١٢): المنحنى المميز للبند (ICC) الافتراضي 1



شكل (٧-١٣): المنحنى المميز للبند (ICC) الافتراضي 2

1أ - بناءً على تلك الأشكال البيانية، قم بتقييم البنود فيما يتعلق بما يلي وشرح ما هي خصائص المنحنى كأساس لتقريرك.

1ب - أي من البندين يقوم بعمل أفضل في النقاط النطاق الكامل للظاهرة التي يقيّمها؟

1ج - أي من خيارات الاستجابة على البند تؤدي وظيفة أفضل في التمييز بين المستويات المختلفة للظاهرة؟

2- صِف الاختلاف في المظهر للمنحنيات المميزة للبنود (ICCs) لبندين ثنائيي الاستجابة binary items مختلفين في معايير التمييز الخاصة بهما.

## الفصل الثامن

### القياس في السياق البحثي الأوسع

يمهد الفصل الافتتاحي لهذا الكتاب المجال لما يجب متابعته من خلال تقديم بعض الأمثلة عن ومتى ولماذا تبرز مشكلات القياس، ومناقشة دور النظرية في القياس، والتأكيد على الاقتصاد الخاطئ جراء العجلة في إجراءات القياس. وبشكل رئيسي، رسم الفصل الافتتاحي السياق البحثي الأوسع قبل أن ينتقل التركيز إلى القضايا المحددة التي تتناولها الفصول اللاحقة. هذا الفصل يعود إلى الصورة الكبيرة وينظر بإيجاز إلى المقياس ضمن السياق الأكبر للبحث.

قبل تطوير المقياس:

ابحث عن الأدوات الحالية:

في بدايات هذا الكتاب، اقترحت أن تطوير المقياس ينشأ غالباً من نقص الأدوات الموجودة المناسبة. إن التأكد من عدم وجود بديل قياس مناسب مسبقاً يعتبر أمراً هاماً وفعالاً. وفي مكان آخر اقترح ديفيليس (DeVellis, 1996) طرقاً للبحث عن مقاييس مناسبة. بشكل أساسي، تتضمن هذه العملية البحث في ملخصات منشورة وإلكترونية للمقاييس لتحديد ما إذا كان هناك مقياس مناسب موجود مسبقاً. وتحتوي السلاسل المنشورة مثل الكتاب السنوي للقياسات الذهنية Mental Measurements Yearbook (مثلاً سبيس، كارلسون، وجينسينجر Spies, Carlson, & Geisinger, 2010) واختبارات في الطباعة (Murphy, Spies, & Plake, 2006) Tests in Print على مقاييس سريرية في المقام الأول، بما في ذلك اختبارات القدرات والشخصية. غالباً ما تكون هذه هي الأدوات التي يستخدمها علماء النفس التطبيقيون لتقييم العملاء. الأدوات المخصصة بشكل أساسي للبحث تكون أقل شهرة، ولكن يتم أخذ بعضها في الاعتبار. هناك فئة أخرى من

المصادر هي المجموعات المستهدفة، مثل المقاييس الكلاسيكية للشخصية Measures of Personality ومقاييس المواقف النفسية الاجتماعية Social Psychological Attitudes (روبينسون، شيفر، فيرجيتسمان، 1991، Robinson, Shaver, & Wrightsman). كما أن المجالات ذات الصلة هي أيضاً مكان ممتاز للعثور على إستراتيجيات القياس التي حققت نجاحاً للآخرين المهتمين بنفس المفهوم البنائي محل الدراسة.

مع التكرار المتزايد، يمكن للباحثين العثور على معلومات حول أدوات القياس على الإنترنت. في الواقع، قد يكون الإنترنت هو المكان الذي يظهر فيه التوسع السريع في المعلومات المتعلقة بالقياس. وكمثال على التوسع في مصادر الإنترنت، أصبح كل من الكتاب السنوي للقياسات الذهنية (Spies et al., 2010) والاختبارات في الطباعة (Murphy et al., 2006) قابلاً للبحث عبر الإنترنت في موقع الويب الخاص بمعهد بيروز Buross للقياسات العقلية في جامعة نبراسكا في لينكولن (<http://www.unl.edu/buross/>). مثال آخر بارز لمصدر من المصادر على الإنترنت للحصول على معلومات حول المقاييس، هو موقع الويب المتعلق بنظام معلومات قياس نتائج تقارير المرضى Patient-Reported Outcomes Measurement Information System (PROMIS). نظام معلومات القياس لنتائج تقارير المرضى PROMIS هو مبادرة بحثية تدعمها المعاهد الوطنية للصحة كجزء من خارطة طريق البحث الطبي (انظر ادر، 2007، Ader)، للحصول على نظرة عامة) لتحقيق ترتيب ووضوح أكبر لتقييم الحالات النهائية المهمة في البحوث الصحية (PROMIS, 2010). هذا الموقع هو بوابة للمعلومات التي تم تطويرها من خلال جهد تعاوني يضم جامعات متعددة وترعاه المعاهد الوطنية للصحة. قام محققو نظام PROMIS بمراجعة وتقييم وتصنيف وتعديل وتطوير بنود تقييم النتائج الصحية في خمسة مجالات واسعة: الأداء البدني، الأداء الاجتماعي، الضيق العاطفي، الألم، والإجهاد. ولقد خضعت هذه البنود للفحص الدقيق باستخدام طرق تستند إلى نظرية الاستجابة للبند IRT. البنود الموجودة في بنوك تلك المجالات المختلفة قد تمثل أكثر البنود التي تمت دراستها بعناية. ومع استمرار جهود نظام PROMIS، ستخضع البنود للتحقق من الصحة على نطاق واسع ومن المحتمل إضافة مجموعات بنود جديدة.

على الرغم من وجود استثناءات، مثل تلك المذكورة أعلاه، تميل مستودعات أدوات القياس القائمة على الإنترنت والمستخدمه على نطاق واسع إلى أن تكون مؤقتة. إن متطلبات التمويل والتوظيف والقضايا الإدارية الأخرى تعقد صيانة مثل هذه المواقع، ونتيجة لذلك، لا يدوم الكثير منها لأكثر من بضع سنوات. ومع ذلك، لا يزال الإنترنت مصدراً قيماً للباحثين الذين يسعون إلى إيجاد المقاييس. تعد محركات البحث على الإنترنت أداة ممتازة لتحديد أدوات القياس التي تستهدف متغيرات معينة. بمعنى أن التطورات في خوارزميات البحث أدت إلى تقليل الحاجة إلى المستودعات الموضوعة على المواقع. فإذا تم تطوير أداة لقياس متغير معين، فمن المرجح أن يجد البحث على الإنترنت إما الأداة نفسها أو معلومات حول كيفية تحديد موقعها.

كما هو الحال مع أي معلومات تستند إلى الإنترنت، يحتاج المستهلك إلى النظر في أصل المعلومات ومصادقيتها. فعادةً ما تقدم المواقع التي ترعاها الجامعات أو الهيئات الحكومية أو المؤسسات أو المنظمات الأخرى المنشأة معلومات دقيقة وذات مصداقية. بشكل عام، ومع ذلك، هناك ما يبرر الحذر. مثلما توجد وفرة من كتب «العلوم التافهة»، هناك أيضاً الكثير من مواقع الإنترنت التي قد تتبنّى أسلوب ومظهر الشرعية العلمية التي لا تتناسب مع محتواها. يجب أن تساعدك المهارات التي اكتسبتها من هذا الكتاب في تقييم مصادر معلومات القياس بشكل ناقد وبأي صيغة وتحديد ما إذا كانت المقاييس الموضحة قد أظهرت ثباتاً وصدقاً كافياً. ويعتبر ما إذا كان المقياس قد تم نشره في مجلة علمية محكمة أم لا، وسيلة أخرى لتقييم موثوقيتها. بالإضافة إلى اجتياز المقياس للتحكيم العلمي، لن يكون مفيداً إلا إذا كان مناسباً لتصور الباحث للمتغير محل الاهتمام. وبالتالي، هناك ما يبرر تمحيص مقياس معين بعد النشر في مجلة ما قبل تقرير أنه يعد اختياراً جيداً لمشروع بحث معين.

عرض المفهوم البنائي في سياق المجتمع محل الاهتمام:

غالبًا ما يكون من المهم تقييم ما إذا كانت المفاهيم البنائية النظرية constructs theoretical التي نحددها نحن الباحثين تتوافق مع التصورات والخبرات الفعلية للأشخاص الذين نخطط لدراساتهم. يمكن أن تكون مجموعات التركيز (مثل Krueger & Casey, 2000) والأساليب النوعية الأخرى (على سبيل المثال، دي والت، روسروك، يونت، ستون 2007, DeWalt, Rothrock, Yount, & Stone, 2007) وسيلة لتحديد ما إذا كانت الأفكار

التي تقوم عليها المفاهيم البنائية محل الاهتمام منطقية بالنسبة للمجيبين. كمثال، خذ بعين الاعتبار السمات - التفسيرات أو التأويلات التي ينسبها الأشخاص إلى نتائج مختلفة. غالبًا ما يتم تقييمها وفقًا لأبعاد مثل "تحت سيطرتي" مقابل "ليس تحت سيطرتي" أو "خاص بهذا الموقف" مقابل "ينطبق على معظم الحالات" أو "خاصيتي" مقابل "خاصية البيئة أو الموقف". كانت البحوث في عمليات معالجة السمات مثمرة للغاية، حيث يمكن لمعظم الناس تحليل النتائج، مثل تلقي عرض عمل بعد المقابلة، على ضوء هذه الأبعاد. في بعض الحالات، قد لا يكون هذا هو الحال. على سبيل المثال، إن مطالبة كبار السن من غير المولودين في المناطق الريفية والمقيمين في المناطق الريفية والذين يفتقرون إلى التعليم، والذين لا يعتادون على التفكير بمثل هذه المصطلحات، بتقييم نتائج المرض أو قرارات الشراء وفقًا لهذه الأبعاد الثلاثة قد لا تعمل بشكل جيد حيث تشير التجربة إلى أنهم قد لا يفهمون المهمة ببساطة لأنها غريبة على طريقة تفكيرهم في الأشياء. إن مجموعة التركيز التي تطلب من المشاركين المحتملين في البحث مناقشة المفاهيم ذات الصلة قد تجعل هذه المشكلة واضحة وتمنع إستراتيجية القياس المحكوم عليها بالفشل.

يمكن لمجموعات التركيز أيضًا أن تكشف عن اللغة الطبيعية اليومية التي يستخدمها الناس للحديث عن مفهوم ما. فقد لا تستخدم الأم الشابة نفس المصطلحات التي يستخدمها خبير التسويق لوصف ردود الفعل على المنتج. وقد تستخدم الأم الشابة «التظاهر» لوصف طفلها يلعب دون استخدام لعبة معينة، بينما قد يصف باحث تسويق مثل هذا الترويج بأنه «منتج غير موجه». بناء البنود بما يتناسب مع لغة الأم المستخدمة (على سبيل المثال، «كم مقدار الوقت من اللعب الذي يقضيه طفلك متظاهراً باللعب، دون استخدام أي ألعاب؟») بدلاً من استخدام لغة الأم الخبيرة (على سبيل المثال، «كم من الوقت يقضيه طفلك في اللعب غير الموجه نحو المنتج؟») من المرجح أن ينتج أداة مناسبة لقياس تصوراتها لكيفية تفاعل طفلها مع مختلف المنتجات.

المقابلات المعرفية هي وسيلة نوعية أخرى لتحديد كيفية تفسير المجيبين المحتملين للبنود وفهمها. على الرغم من وجود مجموعة متنوعة من الأساليب (على سبيل المثال ويليز، Willis, 2005)، فإن الفكرة الأساسية هي التعلم من المجيبين عن طريق سؤالهم عما فهموه حول عنصر ما وكيف صاغوا استجابة له. يمكن أن تكشف هذه الطريقة غالبًا عن التشويش حول المفردات أو المفاهيم أو سوء الفهم المتعلق بخيارات الاستجابة التي قد يتجاهلها الباحث دون إجراء مقابلات

معرفية. يوفر هذا الأسلوب ضمانًا إضافيًا بأن الباحث والمستجيب لديهم فهم مشترك لمعنى البنود، وبالتالي، يمكن أن يعزز ذلك من صدق المقياس.

وكملحوظة تحذيرية: يدعو بعض الباحثين إلى جعل المجتمع المستهدف يعطي الموافقة النهائية على الاستبيانات. هذا أمر مثير للإعجاب ومن المرجح أن يعطي المشاركين إحساسًا أكبر بالمشاركة الفعالة في عملية البحث؛ ومع ذلك، من المهم التعرف على حدود خبرة المجيبين. قد يعرفون أفضل من الباحث كيف يتحدثون هم وأشخاص من خلفيات متشابهة حول قضايا مختلفة، وأنهم مؤهلون بشكل فريد لتقديم رؤى حول فهمهم لمحتويات الاستبيان. يمكن لمجموعات التركيز والمقابلات المعرفية الاستفادة من الخبرة التي يتمتع بها أفراد المجموعة المستهدفة فيما يتعلق بهذه الأمور. لكنهم ليسوا خبراء في التفاصيل الفنية لبناء المقياس. ليس من العدل أن نتوقع من غير الخبراء أن يفهموا القضايا التي تمت مناقشتها في الفصل الخامس، على سبيل المثال، لأنها تنطبق على بناء البنود. وكمثال على ذلك، قد يفضل شخص غير خبير بنداً تمت صياغته بطريقة جذابة ومعتدلة، في حين أن مطور المقياس ذا الخبرة قد يدرك أن هذه الصياغة ستولد تباينًا بسيطًا في الاستجابات، مما يجعل البند غير ذي فائدة. توصيتي الشخصية هي: أن تكون حساسًا لاهتمامات المشاركين، وساعدهم على المشاركة بفعالية في مجموعة متنوعة من الطرق إذا كان ذلك مناسبًا، والاستفادة الكاملة من خبراتهم فيما يتعلق بالجوانب اللغوية أو الثقافية لصياغة البنود، واستمع إلى ملاحظاتهم بشأن المعنى من البنود للمجموعات ذات الاهتمام، ولكن احتفظ بحق الموافقة النهائية على صياغة البند. نحن لا نكرم المشاركين في بحثنا إذا قمنا عن غير قصد بخلق موقف لا يمكن فيه قياس وجهات نظرهم أو مشاعرهم أو آرائهم بدقة؛ نحن ببساطة نضيع وقتهم. الهدف هو فهم ما هي طرق التعبير عن المفاهيم التي ستكون أكثر وضوحًا للمستجيبين مع الحفاظ على سلامة تلك المفاهيم. القيام بذلك يمكن أن يستفيد من فهم الرؤى الخاصة التي يقدمها المشاركون والباحثون والاستفادة من نقاط القوة في مجموعتي المشاركين والباحثين.

اتخاذ القرار بشأن طريقة تطبيق المقياس:

يمكن للباحثين جمع البيانات بعدة طرق (مثل ديلمان، 2007، Dillman، 2007، فاولر، 2009 Fowler)، وقد يختارون طرق تطبيق المقياس بناءً على تفضيلات المجيبين. ووفقًا لذلك، قد يفكر فريق من الباحثين في استخدام طريقة المقابلة بدلاً من طريقة الاستبيان المطبوع.



ومن المهم أن ندرك أن المقياس المراد إكماله في النموذج المطبوع قد يكون له خصائص مختلفة اختلافاً كبيراً عند عرض البنود والردود شفهيًا. على سبيل المثال، قد يكون الآباء أكثر ترددًا في الاعتراف بالطموحات العالية إذا كان عليهم الإبلاغ عنها بصوت عالٍ لمقابلة بدلاً من الاستبيان المطبوع. بشكل عام، من الحكمة قصر أسلوب تطبيق مقياس جديد على الطريقة المستخدمة أثناء تطوير المقياس. وكبدل لذلك، يجب على الباحثين فحص تأثير أسلوب تطبيق المقياس بشكل منتظم. ويمكن استخدام دراسة التعميم (انظر الفصل الثالث) لتحديد مدى إمكانية تعميم المقياس من خلال طرق تطبيق المقياس المختلفة.

الأخذ في الاعتبار المقياس في سياق مقياس أو إجراءات أخرى:

ما هي الأسئلة أو إجراءات البحث التي ستسبق المقياس نفسه؟ كيف ستؤثر هذه الأسئلة على الإجابات على المقياس؟ يشير نونالي (Nunnally, 1978, pp. 627-677)، إلى عوامل سياقية مثل أنماط الاستجابة، والتعب، والدافع كمتغيرات مؤثرة contingent variables. ويشير أيضاً إلى أنه يمكنها التأثير سلباً على الأبحاث من خلال ثلاث طرق: (1) عن طريق تقليل ثبات المقياس؛ (2) من خلال تكوين مصادر موثوقة للتباين بخلاف المفهوم البنائي محل الاهتمام، وبالتالي تقليل الصدق؛ و(3) عن طريق تغيير العلاقات بين المتغيرات، وجعلها تبدو، على سبيل المثال، أكثر ارتباطاً مما هي عليه بالفعل. كمثال على كيفية عمل المتغيرات المؤثرة، فكر في تحفيز المزاج والمجموعات المعرفية لأنها قد تنطبق على مثال أبحاث التسويق. قد يشكل تحفيز المزاج مشكلة إذا قرر الباحثون في التسويق، على سبيل المثال، تضمين مقياس الاكتئاب أو تقدير الذات في نفس الاستبيان كمقياس لتطلعاتهم. غالباً ما تحتوي المقياس التي تنطبق لهذه المفاهيم البنائية (وغيرها) على بنود تعبر عن آراء سلبية عن الذات. إن مقياس روزنبرغ Rosenberg لتقدير الذات (Rosenberg, 1965)، على سبيل المثال، يحتوي على مثل هذه البنود السلبية مثل “أشعر أنه ليس لدي الكثير لأفتخر به” (بالإضافة إلى بنود تعبر عن تقييم ذاتي إيجابي).

يمكن للباحث الذي لم يكن حساساً للآثار المحتملة لاستقراء الحالة المزاجية أن يختار حصرياً سلسلة من البنود ذاتية النقد وضمها إلى مقياس تم تطويره حديثاً. قد تؤدي قراءة العبارات التي تعبر باستمرار عن التقييمات السلبية لنفس الشخص إلى حدوث حالة خلل يمكن أن تتسبب بدورها في إدراك ما هو آت بشكل مختلف عما قد يحدث (على سبيل المثال كيهلستروم، إريك، ساندبراند،

ووتوبايس Tobias & Sandbrand, Eich, Kihlstrom, 2000؛ رهليس، ريسكند، لان، (Rholes, Riskind, & Lane, 1987). هذا قد يؤدي إلى كل من الآثار السلبية الثلاثة التي لاحظها «نونالي» (Nunnally 1978). وهذا يعني أنه في حالة وجود بنود سلبية مؤثرة، قد تأخذ بنود الطموح درجة مختلفة من المعنى، مما يقلل نسبة التباين في تلك البنود التي يمكن عزوها إلى المتغير الكامن المقصود. أو، في الحالة القصوى، قد تتأثر بعض البنود من مقياس الطموح بشكل أساسي بحالة الحالة المزاجية المستحثة، مما يجعل المقياس متعدد العوامل ويخفض صدقه كمقياس لطموح الوالدين. أخيراً، إلى المدى الذي يؤثر فيه مزاج المجيبين على استجاباتهم لبنود الطموح، قد ترتبط درجات هذا المقياس ارتباطاً مصطنعاً بالمقاييس الأخرى ذات الصلة بالمزاج.

المجموعات المعرفية هي مثال أكثر عمومية لنفس الظاهرة. بمعنى أنه قد يتم تحفيز إطار مرجعي ما بخلاف المزاج من خلال تركيز انتباه المجيبين على موضوع محدد. على سبيل المثال، قبل تطبيق مقياس الطموح مباشرة، فإن بنوداً تتعلق بدخل المستجيبين، وقيمة منازلهم، والمبالغ التي ينفقونها سنوياً على فئات مختلفة من السلع الاستهلاكية، قد تخلق حالة ذهنية تعمل على تغيير طموحاتهم تجاه أطفالهم مؤقتاً. نتيجة لذلك، قد تعكس الاستجابات للمقياس حالة عابرة غير مقصودة. كما هو الحال مع الحالة المزاجية، قد تؤثر هذه المجموعة المعرفية سلباً على ثبات و/ أو صدق المقياس عن طريق تلويث المدى الذي يعكس فيه بشكل لا لبس فيه طموحات الوالدين.

#### بعد تطبيق المقياس:

تظهر مجموعة مختلفة تماماً من القضايا بعد استخدام وتطبيق المقياس لمعالجة سؤال بحثي موضوعي. وتبقى الاهتمامات الرئيسية في هذه المرحلة هي تحليل وتفسير البيانات التي تم توليدها من خلال الأداة.

#### قضايا تحليلية:

إحدى القضايا في تحليل البيانات هي مدى ملاءمة التقنيات المختلفة للمتغيرات ذات خصائص القياس المختلفة. لذا؛ فإن المنظور النظري والأساليب التي نوصي بها بقوة في هذا الكتاب يجب أن ينتج عنها مقاييس قابلة لمجموعة واسعة من طرق تحليل البيانات. على الرغم من أن، بالمعنى الحرفي للكلمة، البنود التي تستخدم صيغ ليكرت Likert أو الاستجابة التفاضلية

الدالية semantic differential response قد تكون رتيبة، إلا أن الخبرة الثرية المتركمة تدعم تطبيق الأساليب التحليلية المستندة إلى القياس الفتري interval-based analytic methods على المقاييس التي تنتجها. ومع ذلك، فإن السؤال كان حول ما هي الأساليب الأنسب لنوع البيانات، وبالتأكيد سوف يستمر، في نقاش ساخن في العلوم الاجتماعية. ويعد تحديد كيفية تأثير خيارات الاستجابة المختلفة على تقديرات المتغيرات الأساسية مجاًلاً نشطاً للبحث في حد ذاته. أيضاً، سيكون لدى مختلف الجماهير توقعات مختلفة لكيفية التعامل مع المقاييس. وفي حين أن علماء النفس، على سبيل المثال، قد يكونون متفائلين إلى حد ما في التعامل مع قياسات ليكرت على أنها تنتج بيانات على المستوى الفتري، فإن علماء الأوبئة قد لا يكونون كذلك. ربما يكون النهج الأكثر عملية هو المراقبة (والتوافق) مع التوجه السائد فيما يتعلق بهذه القضية في مجال اهتمام الفرد.

#### قضايا التفسير:

على افتراض أن الباحث قد توصل إلى خطة تحليلية مناسبة للبيانات التي تم إنشاؤها بواسطة مقياس تم تطويره حديثاً، فإن مسألة كيفية تفسير البيانات لا تزال قائمة. إحدى النقاط التي يجب وضعها في الاعتبار عند هذا المنعطف هي أن صدق المقياس لم يتم تأسيسه بشكل قاطع أثناء تطوير المقياس. التحقق من الصدق عملية تراكمية مستمرة. علاوة على ذلك، فإن صدق المقياس هو في الحقيقة إحدى خصائص كيفية استخدام المقياس، وليس المقياس نفسه. مقياس الاكتئاب، على سبيل المثال، قد يكون صادقاً لتقييم الاكتئاب ولكن ليس لتقييم التأثير السلبي العام.

أيضاً، من المهم التفكير في النتائج التي تم الحصول عليها. خاصةً إذا كانت النتائج تبدو غير بديهية أو مضادة للنظرية، فيجب على الباحث عندئذ أن يفكر في احتمال أن يكون المقياس غير صادق في سياق تلك الدراسة بالذات (إن لم يكن على نطاق أوسع). قد يكون مدى صدق المقياس الذي يمكن أن يعمم على المجتمعات أو الأطر أو التفاصيل المحددة لتطبيق المقياس أو مجموعة متنوعة من الأبعاد الأخرى محدوداً. على سبيل المثال، مقياس طموح الوالدين الافتراضي ربما يكون قد تم تطويره مع وجود مجموعة غنية نسبياً في الاعتبار، وقد يكون من غير المقبول صلاحيته للأفراد الذين تكون مواردهم محدودة. يجب أن تأخذ أي استنتاجات تستند إلى مقياس محدود الاستخدام بعين الاعتبار ما يلي: (أ) كيف يختلف تطبيقه الحالي عن سياق الصلاحية

الأصلي الخاص به، (ب) احتمال أن تؤدي هذه الاختلافات إلى الحد من صدق المقياس، و(ج) آثار تلك القيود على البحث الحالي.

#### قابلية التعميم:

تم التحذير في الفقرة السابقة من التعميم على المجتمعات والأطر والجوانب الأخرى للبحث. هذه القضية تستدعي المزيد من التركيز؛ فالتوصل إلى استنتاجات حول الفروق بين المجموعات قد يتداخل مع الاختلافات في الظاهرة التي يتم قياسها والاختلافات في أداء الأداة. إذا استطعنا أن نفترض أن الأخير (أي، الاختلافات في الظاهرة التي يتم قياسها والاختلافات في أداء الأداة) طفيف، فيمكننا أن نسند الفروق الملحوظة إلى عضوية المجموعة. في العديد من المواقف (على سبيل المثال، مقارنة وقت الانتهاء من المهمة في مجموعتين من الأطفال الذين تم اختيارهم وتخصيصهم بشكل عشوائي للمجموعتين)، ستكون هذه هي الحالة. في مجموعات أخرى (على سبيل المثال، المقارنات بين المجموعات المتميزة ثقافيًا)، لا يمكننا افتراض أداء متطابق للأداة. يعد أداء البند التفاضلي differential item functioning، الذي تمت مناقشته باختصار في الفصل السابع، مجالاً نشطاً للبحث السيكميومي. وعلى الرغم من أن معظم الباحثين لن يجعلوا من تلك القضايا محوراً لجهودهم الخاصة، إلا أنه ينبغي عليهم إدراك إمكانية وجودها والقيود التي قد تفرضها على استنتاجاتهم.

#### أفكار ختامية:

#### القياس الصغير والقياس الكبير:

الجانب الأخير من القياس الذي يستدعي النظر هو نطاق المهمة وتعقيد الأدوات اللازمة لها. في مكان آخر من هذا الكتاب (على سبيل المثال، في مناقشة صدق المحتوى)، تحدثت عن مدى اتساع المفهوم البنائي construct الذي صُممت الأداة لتقييمه. وأود العودة هنا إلى موضوع ذي صلة من خلال النظر في كيفية ارتباط نطاق سياق القياس بتعقيد مجمل جهد القياس. هنا سأبدأ بتشبيهه أو محاكاة.

#### الزوارق والسفن البحرية:

عندما كنت طالب دراسات عليا، أمضيت فترة ما بعد الظهر في مساعدة صديق كان يبني زورقاً من مجموعة أدوات اشتراها عبر البريد. لقد كانت هذه مهمة مجدية اقتصادياً وتكنولوجياً لشخص ملتزم بكل النفقات المعتادة والمتطلبات الزمنية لحضور برنامج الدراسات العليا، وقد خدم هذا الزورق صديقي وعائلته جيداً لسنوات عديدة. لو كان قد اقترح عليّ أن أمد له يد العون باستخدام بعض الأموال التي يمكن ادخارها لبناء سفينة بحرية، لربما أحلته إلى مرفق الصحة العقلية بالبحر الجامعي. لكنه لم يكن في حاجة إلى سفينة بحرية لأنواع رحلات المياه التي تصوّرُها وتمتع بها في نهاية المطاف. كان الزورق عملياً ومناسباً للاستخدامات التي قصدُها.

الزوارق والسفن البحرية على حد سواء مراكب قيّمة على الرغم من اختلافها الواضح. ربما يكون الاختلافان الأكثر وضوحاً بينهما حجمهما ووظيفتهما. الزوارق عبارة عن قوارب صغيرة ذات قدرة عالية على المناورة مثالية للاستخدام في المياه المغلقة نسبياً من قبل شخصين أو ثلاثة. إذا كنت ترغب في استكشاف نهر لطيف، واستغراق بعض الوقت لدخول مختلف الخلجان الصغيرة وفحص المواطن الطبيعية بشيء من التفصيل، فإن الزورق المصمم جيداً والمصمم بعناية يعد وسيلة مناسبة للغاية.

سفن الرحلات البحرية هي سفن كبيرة جداً قادرة على استيعاب آلاف الركاب وتزويدهم بمجموعة من وسائل الراحة والأنشطة أثناء عبور البحار المفتوحة. لقد تمت مقارنتها بالمدن العائمة فيما يتعلق بمجموعة الخدمات والمرافق التي تقدمها. وتشمل هذه عادةً قاعات للاحتفالات، ومرافق طعام شاملة، وأماكن إقامة مريحة للنوم، وخدمات متخصصة يقدمها فريق كبير من الموظفين. تعتبر سفينة الرحلات البحرية أكثر من وسيلة مواصلات، فهي تؤدي وظائف متعددة، مثل الترفيه والإقامة، ويمكنها تقديم هذه الوظائف بشكل جيد على قدم المساواة في أجزاء مختلفة من العالم. الحجم والوظيفة ليستا الاختلافات الوحيدة بين هذين النوعين من القوارب المائية. تشغيل كل نوع يتطلب مستويات مختلفة من الخبرة. فمعظم البالغين الأصحاء (والكثير من الأطفال) الذين لديهم بعض الإرشادات والخبرات القصيرة نسبياً يمكنهم تجربة الزورق. من ناحية أخرى، تتطلب سفينة الرحلات البحرية طاقماً مدرباً ومعتمداً من المهندسين البحريين لتشغيلها بأمان. يمثل الوقت اللازم لتصميم الزورق وبناءه وتشغيله مجرد جزء بسيط من النطاق الزمني مقارنةً بجلب سفينة سياحية للخدمة. النماذج التي تتخذها الزوارق وسفن الرحلات البحرية يتم استنباطها من خلال قواعد المعرفة المختلفة، مع توازن مختلف تماماً بين الخبرة والتكنولوجيا المتقدمة التي تميزها.

ربما بسبب بساطتها النسبية، كانت الزوارق بأشكال مختلفة مستخدمة منذ آلاف السنين. وبالتالي، لدينا ثروة من المعرفة العملية عنهم والتي يمكن أن تساعدنا على فهم نقاط القوة والضعف في بدائل التصميم المختلفة. بالمقابل، تتمتع سفن الرحلات البحرية بتاريخ أقصر بكثير ولكنها تستفيد من التطورات التكنولوجية في التصميم والتصنيع. أخيراً، في حين أن تكلفة الزورق متواضعة نسبياً، فإن تكلفة تصميم وبناء وتجهيز وتزويد سفينة الرحلات البحرية أكبر بعدة مرات.

تقوم كل من الزوارق وسفن الرحلات البحرية بعملها بشكل جيد. تعد الزوارق هي الأفضل في حالة الرحلات القصيرة، رغم أنه فيما مضى من الزمن، تم إكمال رحلات لمسافات طويلة بنجاح باستخدام الزوارق. ما نسميه الآن جزر هاواي، على سبيل المثال، من المحتمل أنه تم استعمارها بدايةً من قبل المسافرين من أجزاء أخرى من بولينيزيا حيث كانوا يسافرون مسافات طويلة عبر الزوارق. ومع ذلك، يفضل معظمنا اليوم عبور المياه المفتوحة لجنوب المحيط الهادئ على متن سفينة سياحية، مع الرفاه والراحة والسلامة التي توفرها. على الرغم من أن الزورق كان بإمكانه (ويفترض أن يفعل) إنجاز المهمة، إلا أن سفينة أكثر تطوراً وقدرة تبدو أكثر ملاءمة للوظيفة، والرحلات عبر بولينيزيا في هذه الأيام من المرجح أن تكون باستخدام سفن الرحلات البحرية أكثر من الزوارق. من ناحية أخرى، بمجرد وصول سفينة الرحلات البحرية إلى هاواي من أرخبيل بولينيزي آخر، قد ينتقل المسافر من سفينة الرحلات ويستخدم مركبة أصغر بكثير، مثل الزورق أو قوارب الكاياك، للقيام ببعض الاستكشافات عن قرب في الجزر والممرات المائية الأصغر. لا يمكن ببساطة لسفينة الرحلات الكبيرة المرهقة أن تعمل كزورق عندما يتم تضيق مقاس الممر المائي. أضف إلى ذلك أنه يمكن بسهولة نشر قوارب متعددة لاستكشاف مناطق مختلفة على طول ساحل متعرج أو مجرى مائي دائري.

لذلك أيهما أفضل، سفينة رحلات سياحية أم زورق؟ أود أن أقترح أن لكل منهما مكانه وغرضه، وعلى الرغم من أن وظائفهما تتداخل إلى حد ما، فمن الواضح أن هناك مواقف يفضل أحدهما على الآخر. من المؤكد أن الحصول على الزورق أقل تكلفة، ولكن على الرغم من الرحلات البطولية للزوارق التي استكملت في العصور القديمة، فمن المحتمل أن تكون مقصورة على مهام بسيطة. من ناحية أخرى، فإن التكلفة، والخبرة التكنولوجية، والوقت، والجهد العام الذي ينطوي عليه بناء سفينة سياحية، يحول دون استخدامها في نزاهات مائية أكثر تقييداً تكون فيها مركبة أصغر حجماً كافية تماماً.

## القياس «الزوارق» والقياس «سفن الرحلات البحرية»:

بشكل عام، يمكن للمرء أن يقارن بين «القياس الصغير» و«القياس الكبير» ورسم بعض المقارنات لمناقشتنا للقوارب المائية. ما أشير إليه على أنه قياس صغير ينطوي على تصور وتطوير وتقييم المقاييس التي عادةً ما تكون أكثر تركيزًا وتحديداً. الأمثلة النموذجية هي الحالات العديدة التي يسعى فيها الباحث للإجابة، ضمن مسعى بحثي مقيد، على أسئلة محددة حول بعض المتغيرات التي لا تتوفر لها مقاييس مناسبة. في هذه الحالة، يكرس الكثير من الباحثين الجهود لإنشاء المقاييس التي تتطلبها أبحاثهم. على الرغم من أن الظروف ستختلف بوضوح من مثال إلى مثال للقياس الصغير، فإن الباحث في كثير من الأحيان يتحكم بشكل كبير في كيفية تشغيل البناء على نطاق واسع أو ضيق. يمكن تصميم العناصر التي تم تطويرها لالتقاط البنية بشكل متكرر لاستهداف المتغير الكامن للاهتمام بدقة كبيرة ولتلبية بعض الافتراضات التي يمكن أن تبسط عملية التطوير والتحقق. يمتد التطوير الأولي في كثير من الأحيان لعدة أشهر وليس لعدة سنوات، كما أن الموارد اللازمة والمهام التي ينطوي عليها هذا الجهد يمكن التحكم فيها نسبيًا. القياس الصغير، في هذا الصدد، يشبه الزورق في التشبيه السابق. وهذا يعني أن حجم المشروع بسيط نسبيًا، وستكون وظيفة الأداة المطورة حديثًا محددة تمامًا، ومستوى الخبرة اللازمة لتصميم الأداة في متناول أيدي العديد من الباحثين الأكفاء، والوقت المطلوب في متناول اليد، والإجراء يمكن أن يستفيد من الخبرة العملية بدلاً من الاعتماد على أحدث التطورات، والتكلفة الإجمالية ليست باهظة.

أعني بالقياس الكبير، العمل الذي يهدف إلى تقييم مفاهيم واسعة، ومتعددة الجوانب بطرق تهدف إلى تقديم نتائج ومجموعات أدوات قابلة للتعميم إلى حد كبير ويمكن استخدامها على نطاق واسع عبر مجموعة من الظروف المتنوعة والتي يحتمل أن تكون صعبة. بالإضافة إلى كون المفاهيم نفسها واسعة، فإن مستوى الخبرة المطلوبة لتحقيق قياس كبير حقاً هو في الغالب أمر جوهري. وغالبًا ما تظهر للعيان مثل هذه الأنشطة خلال إطار زمني ممتد، وقد يعتمد تصميم الأدوات بشكل كبير على التطورات المتقدمة في نظرية القياس والممارسة أكثر من الخبرة العملية مع الأساليب المجربة والحقيقية. نتيجة لهذه المطالب المختلفة، فإن التكلفة الإجمالية للقياس الكبير عادةً ما تكون كبيرة.

تشمل الأمثلة على القياس الكبير الاختبارات التعليمية الموحدة مثل اختبار القدرات الدراسية (SAT) واختبارات الخريجين القياسية (GRE) أو مشاريع التقييم الصحي الشاملة مثل مبادرة PROMIS. في هذه الحالات، المتغيرات التي يتم قياسها معقدة وتتضمن مجموعة من المفاهيم الأكثر تحدياً. تشمل القدرات الدراسية، على سبيل المثال، سمات أكثر تمييزاً مثل فهم القراءة والمفردات والكفاءات الرياضية. وبالمثل مبادرة PROMIS واسعة النطاق، تضع مقاييس لمجموعة متنوعة من النتائج المتعلقة بالصحة تتراوح من الاكتئاب إلى المشاركة الاجتماعية إلى قيود التنقل. هذه الأمثلة هي أيضاً هرمية حيث تحدد متغيرات محددة أكثر (مثل المهارات النحوية أو قيود التنقل) كجزء من المفاهيم الأوسع (على سبيل المثال، الاستعداد اللفظي أو الأداء البدني). التقييمات الشاملة للمفاهيم الواسعة مثل هذه النتائج الأكاديمية والصحية تعتبر أنشطة كثيفة الاستخدام للموارد، وعادةً ما تشمل سنوات عديدة من الجهد وفرق البحث الكبيرة ذات المهارات العالية. وكمثال على الخبرة المتخصصة اللازمة للجهود المبذولة في هذا المجال، عندما بدأت مبادرة PROMIS، كانت تتكون من مراكز بحث متعددة تتبع جوانب مختلفة من نطاق النتائج الصحية الأكبر. يضمن كل فريق من هذه الفرق واحداً أو أكثر من المهنيين ذوي الخبرة الكبيرة في القياس. ومع تقدم العمل، واجه باحثو مبادرة PROMIS صعوباتٍ شكلت تحدياً للخبرات المتاحة للفرق الأساسية، وسعى البرنامج إلى الحصول على مدخلات إضافية من خبراء أكثر خبرة وتخصصاً. مع استمرار العمل إلى أبعد من ذلك، استعان الكادر الأصلي من الخبراء الخارجيين بأراء علماء القياس النفسيين الأكثر خبرة وتخصصاً. مشاريع القياس الكبيرة مثل SATs أو PROMIS مماثلةً لسفن الرحلات البحرية متعددة الوظائف التي وصفها سابقاً.

على الرغم من أن أهداف ومتطلبات القياس الصغير عادةً ما تكون أقل من تلك اللازمة للقياس الكبير، فإن الكلمات الوصفية “الكبيرة” و “الصغيرة” لا تهدف إلى الإشارة إلى الاختلافات ذات الأهمية أو القيمة العلمية. في الواقع، العديد من مشاريع القياس الكبيرة مستوحاة من جهود القياس الصغيرة السابقة (تماماً كما قد تتبع سفن الرحلات البحرية اليوم الطرق التي كانت تعبرها الزوارق التي تعود إلى عصور ما قبل التاريخ). يمكن لمقاييس محددة ذات سمات مهمة أن تفرز برامج بحثية عالية الإنتاجية. وفي بعض الحالات، سيتم توسيع نطاقها، وقد تنتقل نماذج القياس المستخدمة من قياس صغير إلى كبير. في حالات أخرى، ستبقى المتغيرات المحددة للغاية محور اهتمام مهم دون أن تصبح أوسع أو تستدعي الانتقال إلى محاولات قياس أكثر تفصيلاً أو واسعة.



## الآثار العملية للقياس الصغير مقابل القياس الكبير:

غالبًا ما تكون النماذج والأدوات التحليلية اللازمة للقياس الصغير أكثر بساطة. بالطبع، يعتمد النموذج «الصحيح» والأسلوب التحليلي «الصحيح» بشكل أساسي على طبيعة أسئلة البحث، ولكن مع القياس الصغير، قد يكون الأمر أكثر ملاءمة لتصميم بنود تلبي افتراضات النماذج الأكثر بساطة والتي تتطلب الأساليب التحليلية الأقل تعقيداً. على سبيل المثال، في الفصل الخامس، عند مناقشة بنود مقياس ليكرت Likert scale، اقترحتُ معايير صياغة البند من خلال تخيل كيف يمكن للمستجيب العادي أن يجيب على كل مجموعة من البنود قيد الدراسة. وأوصيتُ أيضاً بتصميم البنود بحيث يحتمل أن يختار معظم المشاركين استجابة بالقرب من منتصف نطاق خيارات الاستجابة المتاحة لهم. هذا يؤدي أساساً إلى السعي للحصول على بنود ذات مؤشرات حساسة بنفس القدر لقوة مستوى كل مستجيب للمتغير الأساسي. والنتيجة الشائعة لهذه الخطوة البسيطة إلى حد ما هي أنها تميل إلى إنتاج بنود ذات تباينات وتشبعات loadings على العامل المشترك كبيرة نسبياً ومتجانسة إلى حد ما. وهذا يعني أن البنود تميل إلى التوافق بشكل أكبر مع افتراضات مكافئ «تاو» tau equivalence الأساسية. إذا كان الدليل التجريبي يكشف أن التباينات والتشبعات، في الواقع، متجانسة إلى حد ما وأن البنود قد تم إنشاؤها بشكل صريح مع الأخذ في الاعتبار جعلها وبشكل متساوٍ مؤشرات جيدة للمتغير الأساسي، فإن الأساليب القائمة على مبدأ مكافئ «تاو» tau equivalency الأساسي (مثل معامل ألفا) قد تكون مناسبة. (دليلٌ آخر على أن نموذج مكافئ تاو الأساسي tau-equivalent قد يناسب البيانات، هو إنشاء معامل ألفا المحسوب باستخدام إما التباين المشترك أو الصيغ القائمة على الارتباط، تعطي قيمةً متشابهة جداً). أيضاً، محاولة كتابة بنود بحيث أن المستجيب العادي من المرجح أن يختار خيار استجابة بالقرب من مركز النطاق المتاح لخيارات الاستجابة سوف يميل إلى تقليل الالتواء skewness في البيانات. وكما كشفت مناقشة قيود ألفا alpha في الفصل الثالث، فإن تجنب الالتواء في البيانات سوف يخفف من المشاكل المحتملة جراء استخدام ألفا عندما تكون خيارات الاستجابة فعلياً ليست من القياس الفجري interval scaled. وبالمثل، يمكن التبرير بسهولة بإعطاء أوزان متساوية للبنود عندما تكون التباينات والتشبعات متجانسة نسبياً بحيث يمكن لمجموع أو متوسط درجات البنود أن يخدم ببساطة كمخلص للمقياس ككل.

إن التحكم الأكبر الذي يتمتع به الباحث عادةً في حالات القياس الصغير هو أن القياس الأصغر يتيح للباحث أساليب أبسط من الناحية المفاهيمية والتحليلية. قد يكون مفيداً أيضاً، باستخدام نفس البيانات، لمقارنة النتائج بين الأساليب التقليدية (على سبيل المثال، حساب معامل ألفا استناداً إلى ارتباطات بيرسون القياسية) والأساليب المطورة التي تستند إلى افتراضات أكثر ليبرالية (على سبيل المثال، حساب أوميغا كتقدير للثبات reliability). إن المقارنات من هذا النوع، باستخدام البيانات الحقيقية الناتجة من البنود التي تم إنشاؤها مع إدراك افتراضات النموذج الذي صممت البنود لتحقيقه، لا تزال نادرة نسبياً وستقدم خدمة كبيرة لمجتمع البحث.

الأفكار التي نوقشت هنا تتعلق في المقام الأول بثبات أداة القياس، وبطبيعة الحال، فإن مدى ثبات المقياس يفرض قيوداً على صدقه validity. ومع ذلك، كما هو مذكور في الفصل الرابع، يمكن أن يؤثر "التوسع الكبير" في بنود المقياس على الصدق بشكل مباشر (على سبيل المثال، من خلال جعل كل أو بعض محتوى بنود المقياس غير صالح في سياقات معينة). وبالتالي، فإن نطاق جهد القياس المخطط مهم في مراحل متعددة من تطوير المقياس ويستحق النظر فيه بعناية.

إذا كانت البيانات أو أسئلة البحث لا تبرر الأساليب الأبسط لتطوير المقياس، فيجب على الباحث استخدام كل ما هو ضروري لتقديم الأسباب الأوضح والأكثر دقة لعملية القياس والأداة الناتجة. في بعض الحالات، قد يتطلب ذلك البحث عن مشاركة أو مساهمة من متخصصي القياس الأكثر خبرة. لمجرد أن المحيطات كانت قد عُبرت بالزوارق لا يعني أن القيام بذلك في وقتنا الحالي، عندما توجد سفن أكثر قوة، أمر جيد. ولكن على نفس المنوال، إذا أمكن تبسيط العملية بما يحقق الأهداف البحثية المقصودة بالقدر الكافي، فإنه يمكن أن تكون البساطة الناتجة ذات فائدة كبيرة. بالإضافة إلى ذلك، الجهود والموارد غير الضرورية التي لا تفيد مرحلة القياس قد يتم تحويلها لصالح بعض الجوانب الأخرى للمشروع. وعلى الرغم من وجود قيمة في فهم كيفية توظيف نماذج القياس والأساليب التحليلية الأكثر تعقيداً، إلا أن هناك قيمة مكافئة في إدراك - ويفضل أن يكون ذلك في مرحلة مبكرة من العمل - أن توظيفها غير ضروري حقاً لتحقيق الهدف المطروح. غالباً ما يوفر القياس الصغير فرصاً لمثل هذا التبسيط الذي نادراً ما يحدث في القياس الكبير.

تذكّر، القياس مسألة هامة:

القياس يمثل جانباً حيويًا في البحوث الاجتماعية والسلوكية. بغض النظر عن مدى جودة تصميم وتنفيذ جوانب أخرى من المسعى البحثي، يمكن للقياس بناء أو هدم الدراسة. نفترض أن المتغيرات التي تهتمنا تتوافق مع إجراءات التقييم التي نستخدمها. في كثير من الأحيان، العلاقة محل الاهتمام الرئيسي تكون بين اثنين أو أكثر من المتغيرات التي لا يمكن ملاحظتها أو رصدها، مثل الرغبة في الحصول على نتيجة معينة والفشل في التفكير في نتائج بديلة. لا يمكننا قياس الرغبة أو الاعتبار مباشرة، لذلك نبني مقاييس نأمل أن تتضمنها. هذه المقاييس، هي إلى حد ما، استعارات كمية للمفاهيم الأساسية. فقط إلى الحد الذي تكون فيه هذه الاستعارات ملائمة (أي الأدوات صالحة)، فإن العلاقات التي نلاحظها بين المقاييس تعكس العلاقة التي نرغب في تقييمها بين المفاهيم البنائية غير القابلة للملاحظة أو الرصد. أخذ العينات الممثلة، وتصميم البحث المميز، والتنفيذ المتقن للإجراءات لن يغير هذه الحقيقة. الباحث الذي لا يفهم العلاقة بين المقاييس والمتغيرات التي تمثلها، بالمعنى الحرفي، لا يعرف ما الذي يتحدث عنه. إذا نظرنا إليها في ضوء ذلك، فإن الجهود التي ينطوي عليها القياس الدقيق تتم مكافأتها بشكل وافٍ من خلال فوائدها.

#### تمارين:

1- صف بإيجاز القياس الصغير والقياس الكبير، واذكر مزايا وعيوب كل منهما، وقدم أمثلة بحثية لكل منهما.

## المراجع

- Ader, D. N. (2007). Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical Care, 45 (5, Suppl. 1), S1-S2.
- Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting behavior. Englewood Cliffs, NJ: Prentice Hall.
- Alder, K. (2002). The measure of all things: The seven-year odyssey world. New York: Free Press. and hidden error that transformed the
- Allen, M. J., & Yen, W. M. (1979). Introduction to measurement theory. Monterey, CA: Brooks/Cole.
- American Psychiatric Association. (1952). Diagnostic and statistical Author. :manual: Mental disorders. Washington, DC
- American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text revision). Washington, DC: Author.
- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author.

- Anastasi, A. (1968). *Psychological testing* (3rd ed.). New York: Macmillan.
- Aragona, M. (2009). The concept of mental disorder and the DSM-V. *Dialogues in Philosophy, Mental, and Neuro Sciences*, 2, 1-14.
- Asher, H. B. (1983). *Causal modeling* (2nd ed.). Sage university paper series on quantitative applications in the social sciences (Series No. 07-003). Beverly Hills, CA: Sage.
- Barnette, W. L. (1976). *Readings in psychological tests and measurements* (3rd ed.). Baltimore: Williams & Wilkins.
- Bjorner, J. B., Chang, C., Thissen, D., & Reeve, B. B. (2007). Developing tailored instruments: Item banking and computerized adaptive assessment. *Quality of Life Research*, 16, 95-108.
- Blalock, S. J., DeVellis, R. F., Brown, G. K., & Wallston, K. A. (1989). Validity of the Center for Epidemiological Studies Depression Scale in arthritis populations. *Arthritis and Rheumatism*, 32, 991-997.
- Bohrnstedt, O. W. (1969). A quick method for determining the reliability and validity of multiple-item scales. *American Sociological Review*, 34, 542-548.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Buchwald, J. Z. (2006). Discrepant measurements and experimental knowledge in the early modern era. *Archive for History of Exact Sciences*, 60, 565-649.

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carmines, E. G., & McIver, J. P. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. W. Bohrnstedt & B. F. Borgatta (Eds.), *Social measurement: Current issues* (pp. 65-115). Beverly Hills, CA: Sage.
- Cattell, R. B. (1966). The screen test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Choi, S. W., Reise, S. P., Pilkonis, P. A., Hays, R. D., & Cella, D. (2010). Efficacy of static and computer adaptive short forms compared to full-length measure of depressive symptoms. *Quality of Life Research*, 19, 125-136.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46.
- Cohen, P., Cohen, J., Teresi, J., Marchi, M., & Velez, C. N. (1990). Problems in the measurement of latent variables in structural equation causal models. *Applied Psychological Measurement*, 14, 183-196.
- Comrey, A. L. (1973). *A first course in factor analysis*. New York: Academic Press.
- Comrey, A. L. (1988). Factor analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology*, 56, 754-761.

- Converse, J. M., & Presser, S. (1986). Survey questions: Handcrafting the standardized questionnaire. Sage university paper series on quantitative applications in the social sciences (Series No. 07-063). Beverly Hills, CA: Sage.
- Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart & Winston.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). Dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cureton, E. E. (1983). Factor analysis: An applied approach. Hillsdale, NJ: Lawrence Erlbaum.
- Currey, S. S., Callahan, L. F., & DeVellis, R. F. (2002). Five-item Rheumatology Attitudes Index (RAI): Disadvantages of a single positively worded item. Unpublished paper, Thurston Arthritis Research Center, University of North Carolina at Chapel Hill.
- Czaja, R., & Blair, J. (1996). Designing surveys: A guide to decisions and procedures. Thousand Oaks, CA: Pine Forge.
- :- Dale, F., & Chall, J. E. (1948). A formula for predicting readability Instructions. *Education Research Bulletin*, 27, 37-54.

- Davey, Z. Z. T., & Hendrickson, A. (2010, May). Classical versus IRT statistical test specifications for building test forms. Paper presented at the annual meeting of the National Council of Measurement Education, Denver, Colorado.

- De Boeck, P., & Wilson, M. (2004). Explanatory item response models: A generalized linear and nonlinear approach. New York: Springer.

- Decker, H. S. (2007). How Kraepelinian was Kraepelin? How Kraepelinian are the neo-Kraepelinians? From Emil Kraepelin to DSM-III. *History of Psychiatry*, 18(3), 337-360.

s guide to finding, evaluating,' - DeVellis, R. F. (1996). A consumer and reporting on measurement instruments. *Arthritis Care and Research*, 9, 239-245.

- DeVellis, R. F. (2005). Inter-rater reliability. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (Vol. 2, pp. 317-322). San Diego: Elsevier.

- DeVellis, R. F., & Callahan, L. F. (1993). A brief measure of helplessness: The helplessness subscale of the Rheumatology Attitudes Index. *Journal of Rheumatology*, 20, 866-869.

- DeVellis, R. F., DeVellis, B. M., Blanchard, L. W., Klotz, M. L., Luchok, K., & Voyce, C. (1993). Development and validation of the Parent Health Locus of Control (PHLOC) scales. *Health Education Quarterly*, 20, 211-225.



- DeVellis, R. F., DeVellis, B. M., Revicki, D. A., Lurie, S. J., Runyan, D. K., & Bristol, M. M. (1985). Development and validation of the child improvement locus of control (CILC) scales. *Journal of Social and Clinical Psychology*, 3, 307-324.
- DeVellis, R. F., Holt, K., Renner, B. R., Blalock, S. J., Blanchard, L. W., Cook, H. L., et al. (1990). The relationship of social comparison to rheumatoid arthritis symptoms and affect. *Basic and Applied Social Psychology*, 11, 1-18.
- DeWalt, D. A., Rothrock, N., Yount, S., & Stone, A. A. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care*, 45(5, Suppl. 1), S12-S21.
- Dillman, D. A. (2007). *Mail and Internet surveys: The tailored design* (2nd ed., 2007 update). Hoboken, NJ: Wiley.
- Duncan, O. D. (1984). *Notes on social measurement: Historical and critical*. New York: Russell Sage.
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to solution to the pervasive problem of internal omega: A practical consistency estimation. *British Journal of Psychology*, 105, 399-412.
- Embretson, S. E., & Hershberger, S. L. (1999). Summary and future of psychometric models in testing. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement* (pp. 243-254).
- Mahwah, NJ: Lawrence Erlbaum. Embretson, S. E., & Reise, S. P. (2010). *Item response theory* (2nd ed.). New York: Routledge Academic.

- Eysenck, H. J. (1968). Kraepelin, Emil. International Encyclopedia of the Social Sciences. Retrieved October 9, 2015, from <http://www.encyclopedia.com/doc/1G2-3045000662.html>.
- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58, 357-381.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
- Fink, A. (1995). *The survey kit*. Thousand Oaks, CA: Sage.
- Fowler, F. J., Jr. (2009). *Survey research methods* (4th ed.). Thousand Oaks, CA: Sage.
- s readability graph: Clarifications, validity, and' - Fry, E. (1977). Fry extension to level 17. *Journal of Reading*, 21, 249.
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research and Evaluation*, 17, 1-13.
- Ghiselli, B. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco: Freeman.
- Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.

- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991).  
Fundamentals of item response theory. Newbury Park, CA: Sage.
  - Harman, H. H. (1976). Modern factor analysis. Chicago: University  
of Chicago Press.
  - Hathaway, S. R., & McKinley, J. C. (1967). Minnesota Multiphasic  
Personality Inventory: Manual for administration and scoring. New York:  
Psychological Corporation.
  - Hathaway, S. R., & Meehl, P. E. (1951). An atlas for the clinical use  
of the MMPI. Minneapolis: University of Minnesota Press.
  - Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention  
decisions in exploratory factor analysis: A tutorial on parallel analysis.  
Organizational Research Methods, 7(2), 191-205.
  - Hewett, J. E., Anderson, S. K., & Minor, M. A. (1992). Statistical  
analyses for research in arthritis when assessing change. Arthritis Care and  
Research, 5, 74-80.
  - Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality:  
A review of twenty-seven community studies. Journal of Health and Social  
Behavior, 38, 21-37.
- s blog: Transforming diagnosis.'- Insel, T. R. (2013). Director  
Accessed July 21, 2015, from  
[www.nimh.nih.gov/about/director/2013/transforming-//:http  
diagnosis.shtml](http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml).

- Insel, T. R. (2015). Dr. Tom Insel to step down as NIMH director. Accessed October 9, 2015, from <http://www.nimh.nih.gov/about/drtom-insel-to-step-down-as-nimh-director.shtml>.
- Insel, T. R., & Leiberian, J. A. (2013). DSM-5 and RDoC: Shared interests. Retrieved September 18, 2015, from [www.nimh.nih.gov/news/science-news/2013/dsm-5-and-rdoc-shared-interests.shtml](http://www.nimh.nih.gov/news/science-news/2013/dsm-5-and-rdoc-shared-interests.shtml).
- Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 109-134.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Keefe, F. J. (2000). Self-report of pain: Issues and opportunities. In A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 317-337). Mahwah, NJ: Lawrence Erlbaum.
- Kelly, J. R., & McGrath, J. B. (1988). *On time and method*. Newbury Park, CA: Sage.
- Kihlstrom, J. F., Eich, E., Sandbrand, D., & Tobias, B. A. (2000). Emotion and memory: Implications for self-report. In A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 81-103). Mahwah, NJ: Lawrence Erlbaum.
- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). San Francisco: Brooks/Cole.

- Kraepelin, E. (1883). *Compendium der Psychiatrie zum Gebrauche Studirende und Aerzte*. Leipzig: Abel. für
- Krueger, R. A., & Casey, M. A. (2000). *Focus groups: A practical guide for applied research*. Thousand Oaks, CA: Sage.
- Kreuger, R. F., Watson, D., & Barlow, D. H. (2005). Introduction to the special section: Toward a dimensionally based taxonomy of psychopathology. *Journal of Abnormal Psychology*, 114, 491-493.
- Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2014). Assessing individual change using short tests and questionnaires. *Applied Psychological Measurement*, 38, 201-216.
- Levenson, H. (1973). Multidimensional locus of control in psychiatric patients. *Journal of Consulting and Clinical Psychology*, 41, 397-404.
- Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and post-testing periods. *Review of Educational Research*, 47, 121-150.
- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.
- Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis*. Mahwah, NJ: Lawrence Erlbaum.
- Long, J. S. (1983). *Confirmatory factor analysis*. Sage university paper series on quantitative applications in the social sciences (Series No. 07-033). Beverly Hills, CA: Sage.

- Lord, F. M., & Novick, M. R. (2008). Statistical theories of mental test scores.

Charlotte, NC: Information Age.

- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.

- Mayer, J. M. (1978). Assessment of depression. In M. P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 4, pp. 358-425). San Francisco: Jossey-Bass.

- McDonald, R. P. (1984). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum.

- Meehl, P. E. (1999). Clarifications about taxometric method. *Applied & Preventive Psychology*, 8, 165-174.

- Messick, S. (1995). Validity of psychological assessment: Validation responses and performances as scientific 'of inferences from persons inquiry into score meaning. *American Psychologist*, 50, 741-749.

- Mitchell, S. K. (1979). Interobserver agreement, reliability, and generalizability of data collected in observational studies. *Psychological Bulletin*, 86, 376-390.

s walk: How randomness rules'- Mlodinow, L. (2008). *The drunkard our lives*. New York: Pantheon.

- Murphy, L. L., Spies, R. A., & Plake, B. S. (2006). *Tests in print*. Lincoln: University of Nebraska Press.

3rd ed.).) - Myers, J. L. (1979). Fundamentals of experimental design  
Boston: Allyn & Bacon.

- Namboodiri, K. (1984). Matrix algebra: An introduction. Sage  
university paper series on quantitative applications in the social sciences  
(Series No. 07-028). Beverly Hills, CA: Sage.

- Narens, L., & Luce, R. D. (1986). Measurement: The theory of  
numerical assignments. *Psychological Bulletin*, 99, 166-180.

- Nering, M. L., & Ostini, R. (2010). Handbook of polytomous item  
response theory models. New York: Routledge.

- Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York:  
McGraw-Hill.

- Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd  
ed.). New York: McGraw-Hill.

- Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of  
of attitude change. *Psychological Bulletin*, 62, congruence in the prediction  
42-55.

- Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference  
scores: A paradox for measurement of change. *Psychological Bulletin*,  
82(1), 85-86.

- Padilla, M. A., Divers, J., & Newton, M. (2013). Coefficient alpha  
bootstrap confidence interval under normality. *Applied Psychological  
Measurement*, 36, 331-348.

- PBS. (2002). Timeline: Treatments for mental illness. Retrieved July 21, 2015, from <http://www.pbs.org/wgbh/amex/nash/timeline>.
- Peters, G. (2014). The alpha and the omega of scale reliability and Cronbach's alpha and the route towards validity: Why and how to abandon Cronbach's more comprehensive assessment of scale quality. *The European Health Psychologist*, 16, 56-69.
- PROMIS. (2010). Patient-Reported Outcomes Measurement Information System: Dynamic tools to measure health outcomes from the patient perspective.
- Retrieved from <http://www.nihpromis.org/default.aspx>
- Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Chicago: MESA.
- 'Top mental health researcher joined Alphabet'. Regalado, A. (2015, September 21). Why America's top mental health researcher joined Alphabet. Retrieved October 15, 2015, from <http://www.technologyreview.com/news/541446/why-americas-topmental-health-researcher-joined-alphabet>.
- Reise, S. P., Moore T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92, 544-559.



- Reise, S. P., & Revicki, D. A. (2015). Handbook of item response theory modeling: Applications to typical performance assessment. New York: Routledge.
- Rholes, W. S., Riskind, J. H., & Lane, J. W. (1987). Emotional states and memory biases: Effects of cognitive priming and mood. *Journal of Personality and Social Psychology*, 52, 91-99.
- Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). Measures of personality and social psychological attitudes. San Diego: Academic Press.
- Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171-210). New York: Springer.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Rotter, J. B. (1966). Generalized expectancies for internal vs. external control of reinforcement. *Psychological Monographs*, 80(1, Whole No. 609).
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph* (Suppl. 17).
- Saucier, G., & Goldberg, L. R. (1996). The language of personality: five-factor model. In J. S. Wiggins (Ed.), *The Lexical perspectives on the five-factor model of personality* (pp. 21-50). New York: Guilford.

- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited  $\alpha$ . *Psychometrika*, 74, 107-120.'usefulness of Cronbach
- Silvestro-Tipay, J. L. (2009). Item response theory and classical test theory: An empirical comparison of item/person statistics in a biological science test. *International Journal of Educational and Psychological Assessment*, 1, 19-31.
- Smith, P. H., Earp, J. A., & DeVellis, R. F. (1995). Measuring s Experiences with Battering (WEB)'battering: Development of the Women s Health: Research on Gender, Behavior, and Policy, 1, 273-'scale. *Women* 288.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). State-trait anxiety inventory (STAI) test manual for form X. Palo Alto, CA: Consulting Psychologists Press.
- Spies, R. A., Carlson, J. F., & Geisinger, K. F. (2010). The eighteenth mental measurements yearbook. Lincoln: University of Nebraska Press.
- Stage, C. (2003). Classical test theory or item response theory: The Swedish experience (EM No. 42). Umea, Sweden: Umea Universitet Department of Educational Measurement. Available online at [http://www8.umu.se/edmeas/publikationer/index\\_eng.html](http://www8.umu.se/edmeas/publikationer/index_eng.html)

- Sterba, K. R., DeVellis, R. F., Lewis, M. A., Baucom, D. H., Jordan, J. M., & DeVellis, B. M. (2007). Developing and testing a measure of dyadic efficacy for married women with rheumatoid arthritis and their spouses. *Arthritis & Rheumatism (Arthritis Care & Research)*, 57(2), 294-302.
  
- Strahan, R., & Gerbasi, K. (1972). Short, homogenous version of the of Clinical Psychology, Marlowe-Crowne Social Desirability Scale. *Journal* 28, 191-193.
  
- Tartakovsky, M. (2011). How the DSM developed: What you might not know. Retrieved October 9, 2015, from <http://psychcentral.com/blog/archives/2011/07/02/how-the-dsm-developed-whatyou-might-not-know>.
  
- Thissen, D., Chen, W.-H., & Bock, R. D. (2003). Multilog (Version 7.0) [Computer software]. Lincolnwood, IL: Scientific Software International.
  
- Timmermann, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16, 209-220.
  
- Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, 34, 414-424.
  
- Traub, R. E. (1994). Reliability for the social sciences: Theory and applications.

- Thousand Oaks, CA: Sage.
- Van der Linden, W. J., & Glas, C. A. W. (2000). Computerized adaptive testing: Theory and practice. St. Paul, MN: Assessment Systems.
- Venables, W. N., Smith, D. M., & the R-Core team. (2015). An introduction to R. Notes on R: A programming environment for data analysis and graphics. Available for download from <https://cran.r-project.org/manuals.html>.
- Wallston, K. A., Stein, M. J., & Smith, C. A. (1994). Form C of the MHLC Scales: A condition-specific measure of locus of control. *Journal of Personality Assessment*, 63, 534-553.
- Wallston, K. A., Wallston, B. S., & DeVellis, R. (1978). Development and validation of the multidimensional health locus of control (MHLC) scales. *Health Education Monographs*, 6, 161-170.
- Weisberg, H., Krosnick, J. A., & Bowen, B. D. (1996). An analysis. Thousand Oaks, introduction to survey research, polling, and data CA: Sage.
- Willis, G. B. (2005). Cognitive interviewing: A tool for improving questionnaire design. Thousand Oaks, CA: Sage.
- Wright, B. D. (1999). Fundamental measurement for psychology. In S. L. Hershberger (Eds.), *The new rules of measurement* & S. E. Embretson (pp. 65-104). Mahwah, NJ: Lawrence Erlbaum.
- Yu, C. H. (2005). Test-retest reliability. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (Vol. 3, pp. 777-784). San Diego, CA:

Elsevier.

- Zickar, M. J., & Broadfoot, A. A. (2008). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends*. New York: Routledge Academic.
- Zorzi, M., Priftis, K., & Umilita, C. (2002, May). Brain damage: Neglect disrupts the mental number line. *Nature*, 417, 138-139.
- Zuckerman, M. (1983). The distinction between trait and state scales is not arbitrary: Comment on Allen and Potkay's "On the arbitrary distinction between traits and states." *Journal of Personality and Social Psychology*, 44, 1083-1086.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

## Notes

[1←]

مجموعة الأعداد الطبيعية تبدأ من الصفر إلى موجب ما لا نهاية بزيادة واحد صحيح في كل مرة. أما مجموعة الأعداد الصحيحة، فتشتمل على الأعداد من سالب ما لا نهاية بالإضافة إلى الصفر بالإضافة إلى الأعداد الموجبة التي تحتويها مجموعة الأعداد الطبيعية بزيادة واحد صحيح كل مرة. أما الأعداد الكسرية، فتتكون من كسور الأعداد الصحيحة في صورة بسط ومقام. أما الأعداد الحقيقية، فتشمل المجموعات السابقة كلها بالإضافة إلى الأعداد التي لا يمكن كتابتها على شكل كسور مثل باي ( $\pi$ ) (أي الأعداد اللاكسرية). ويمكن تصور الأعداد الحقيقية بأنها أعداد غير متناهية على خط مستقيم. وتأخذ الأعداد الحقيقية اسمها من تضادها مع فكرة الأعداد التخيلية. كما يمكن لها أن تقوم بقياس الكميات المستمرة على اختلافها (المترجمين).

[2←]

الدرجة الحقيقية هي مفهوم افتراضي، يشير إلى خصائص لا يمكن ملاحظتها، ولا يمكن قياسها بشكل مباشر. وتُعرف إجرائياً بأنها متوسط درجات الفرد المتوقعة على سمة أو متغير معين عند قياس السمة أو المتغير عدداً لا نهائياً من المرات. وبسبب عدم إمكانية إعادة تطبيق الاختبار على الفرد إلى ما لا نهاية فإن إمكانية التحديد التام للدرجة الحقيقية يعد أمراً تقريبياً (المترجمين).

[3←]

بوتستراب (Bootstrap) هي تقنية لإعادة المعاينة والتي هي حالة خاصة من أساليب المحاكاة بطريقة مونت كارلو وذلك للحصول على مقدرات أكثر دقة عندما يكون حجم العينة صغراً، أو تعذر تحديد التوزيع الاحتمالي للمقدر محل الاهتمام. والفكرة الأساسية لهذه التقنية تتضمن القيام بتوليد عدد كبير من العينات المسحوبة مع الإرجاع من بيانات العينة الأصلية. كذلك، فهذه الطريقة تزودنا بالخطأ المعياري للمقدرات بدون أن يتطلب ذلك حسابات نظرية، وهذا يوفر الوقت والجهد (المترجمين).